3′ UTR SEQUENCES AND SYNTAX: INVESTIGATING *CIS*-ELEMENT DENSITY AND INTERACTIONS, AND THE ROLE OF 3′ UTR LENGTH IN GENE REGULATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Katla Kristjánsdóttir

January 2017

3′ UTR SEQUENCES AND SYNTAX: INVESTIGATING *CIS*-ELEMENT DENSITY AND INTERACTIONS, AND THE ROLE OF 3′ UTR LENGTH IN GENE REGULATION

Katla Kristjánsdóttir, Ph.D.

Cornell University 2017

Regulating the precise rate of protein production from each protein-coding gene is a fundamental process of all cellular life. While transcriptional regulation plays a large role in determining final protein levels, post-transcriptional events can also make substantial contributions. In mammals, the majority of the *cis*-regulatory information that controls post-transcriptional events is located within a transcripts 3′ untranslated region (3′ UTR). The *cis*-regulatory sequence elements (*cis*-elements) found within 3′ UTRs are bound by *trans*-acting factors, mainly RNA binding proteins and non-coding RNAs, which in turn interact with the core decay and translation machineries to modulate mRNA decay or protein synthesis rates. Though a large number of *cis*-elements have been identified, many questions remain about their distribution and interactions. In addition, the contribution of parameters whose function is independent of their sequence, such as the length of the 3′ UTR, to gene regulation is poorly understood.

Numerous studies have established that typical 3′ UTRs contain multiple discrete *cis*-elements, yet the typical density of elements within 3′ UTRs is unclear. Moreover, examples exist describing consequential interactions between *cis*-elements, either cooperative or inhibitory. However, the extent to which interactions are a general paradigm for *cis*-elements remains to be determined. By performing a systematic study of the regulatory sequence information within two conserved mammalian 3′ UTRs, those of *Hmga2* and *PIM1*, I determined that both 3′ UTRs contain a high density of *cis*-elements (at minimum 6 and 12 per kb, respectively) spread across the entire 3′UTR. Importantly, the vast majority of the *cis*-elements function independently of neighboring elements. Addition-

ally, despite the overall repressive effect of the 3′ UTRs, I found that many regulatory *cis*-elements enhance gene expression, rather than repressomg it. I hypothesize that the enhancing *cis*-elements counteract a repressive effect of 3′ UTR length.

In a second study, I explored the effect of 3′ UTR length on gene expression using, as 3'UTR mimics, randomly-generated, nucleotide-composition matched, sequences of varying lengths. Long 3′ UTRs have previously been identified as targets of an mRNA surveillance mechanism called nonsense-mediated decay (NMD). In this study, I discovered a novel role for 3′ UTR length in triggering an NMD-independent decay pathway in human cell lines. Reporter transcripts with random 3′ UTR mimics as short as 400 nucleotides were repressed by this pathway, with the repression growing stronger with increasing length. While the mechanism of this novel pathway remains to be elucidated, I have determined that it affects the decay rate of mature mRNAs in a deadenylation-independent manner.

Overall, by determining the density and extent of interactions of *cis*-element within example mammalian 3′ UTRs and by identifying a novel role for 3′ UTR length in regulating gene expression, this work furthers our understanding of fundamental aspects of 3′ UTR-mediated gene regulation.

BIOGRAPHICAL SKETCH

Katla Kristjánsdóttir was born in Denmark in 1987 while her parents were studying at the University of Copenhagen. She moved to her native Iceland two years later. In school, she enjoyed most subjects though language classes were always her favorite. Through most of high school, she planned on becoming an interpreter for the UN. However, both chemistry and biology caught her interest and by the time she had to choose a subject to study at University, she had taken a particular fancy to studying the inner workings of eukaryotic cells.

Katla enrolled in Biochemistry at the University of Iceland in 2007. While there, she completed two different summer research programs and a B.S. thesis research project. In the summer of 2008, she worked in Dr. Gudmundur H. Gudmunssons laboratory, studying how the expression of the antimicrobial peptide LL37 is induced. The following summer, she had the opportunity to join an exchange program that took her to Caltech in Pasadena, CA to work in the laboratory of Dr. Judith Campbell. There she studied the yeast replication checkpoint protein, Mrc1, and identified its preferred DNA binding substrates. When she returned to the University of Iceland, she joined the laboratory of Dr. Eirkur Steingrimsson to do her B.S. thesis work on the regulation of an oncogenic transcription factor, MITF. Upon graduating from the University of Iceland in 2010, she received the Gudmundur P. Bjarnason prize for outstanding performance in a biochemistry degree program. Before heading off to graduate school, she worked as a research technician in the laboratory of Dr. Stefan Sigurdsson, studying the ATPase activity of the DNA repair protein Rad26.

In August of 2010, Katla joined the field of Biochemistry, Molecular and Cell Biology at Cornell University. She joined the laboratory of Dr. Andrew Grimson, to study post-transcriptional gene regulation.

This dissertation is dedicated to my husband, whose unwavering support and encouragement has been invaluable.

ACKNOWLEDGEMENTS

There are many people I would like to thank for helping me on this journey. First, I would like to thank my advisor, Andrew Grimson, for accepting me into his research group and for being an excellent mentor to me throughout the years. I would also like to thank all past and current members of the Grimson lab for contributing to the wonderful environment in the Grimson lab and providing invaluable help and feedback, interesting discussions (scientific and non-scientific), and being a source of emotional support. I am particularly grateful to Erin Wissink, who was my mentor when I rotated and continued to be a source of information, wisdom, and friendship throughout our time together in the lab. I would also like to warmly thank Elizabeth Fogarty, our lab manager, with whom I have collaborated closely for years. Apart from being a first-class lab manager, she is a dedicated and hard-working collaborator and a dear friend.

There are many in MBG I would like to thank. The Pleiss lab, which has provided insightful ideas and critiques during the joint Grimson-Pleiss lab meetings and journal clubs. My committee members for providing excellent suggestions and feedback. The labs I rotated in, the Smolka lab and the Soloway lab, that gave me a chance to learn so many interesting things. Our graduate field advisers, who have provided patient assistance with University paperwork. Everyone who showed up for my field seminars and asked insightful questions; particularly, Volker Vogt, who sent me helpful feedback on my presentation afterwards. Jen Grenier, who has been very helpful in designing and troubleshooting qPCR and shRNA KD experiments. Eric Alani, whose cheerful presence on the fourth floor is always a positive contribution to my day. Bill Brown, as well as Maki Inada at Ithaca College, who let me guest lecture in their classes and provided me with feedback on my performance. And finally, all the Biotech building staff for keeping Biotech running and for helping me when I have lost things in the building.

During my first semester I had health problems and was admitted to hospital twice. MBG was incredibly supportive and made me feel less alone while hospitalized in a new

country. Marcus Smolka was patient and supportive as I missed many days in lab. Volker Vogt, who was DGS at the time, came to visit me in the hospital. Linda Nicholson sent me flowers on behalf of the class I was taking with her that semester. My fellow BMCB students supported me by visiting me in the hospital, by listening when I was upset, and by providing notes for classes I missed. I am particularly grateful to Ian and Carolyn who drove me to the ER on several occasions, as I did not have a car at the time.

Finally, endless thanks go to my friends and family. My fellow BMCB students were wonderful friends and a strong support system, which was especially valuable during the first couple of years as I adjusted to a new country. My childhood friends in Iceland always believe in me and their love and support carries me onward. I am fortunate in belonging to a wonderful family. My loving parents are both excellent role models and raised me to work hard and to believe in the value of education. My genius little brother keeps in touch and shows interest in my life and work. Finally, I thank my husband with all my heart. He transferred Universities so that we could be together. He learned swing dancing because he knew I loved to dance. He has been unshakably loving and supportive, giving our relationship first priority, and doing everything he can to make my life easier and more wonderful.

The work presented in this dissertation was supported by a Research Scholar Grant from the American Cancer Society (RSG-13-057-01-RMC), and National Institutes of Health (NIH) grant 1R01GM105668.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADAR	Adenosine deaminase acting on RNA
Ago	Argonaute
APA	Alternative polyadenylation
ARE	AU-rich element
ARE-BP	ARE-binding protein
CARE	CA-rich element
DNA	Deoxyribonucleic acid
dsRBP	Double-stranded RNA-binding protein
dsRNA	Double-stranded RNA
ER	Endoplasmic reticulum
FACS	Fluorescence activated cell sorting
Fig.	Figure
FL	Full Length
GRE	GU-rich element
m5C	5-methylcytosine
m6A	N6-methyladenosine
m7	*Hmga2* 3′ UTR reporter with mutated *let-7* target sites
miRISC	miRNA-induced silencing complex
mRBP	mRNA bound RBP
mRNP	Messenger-ribonucleoproteins
NMD	Nonsense-mediated decay
NP-SD	Non-parametric estimate of standard deviation
nt	Nucleotide
nts	Nucelotides
P/S	Penicillin/Streptomycin
PAS	Polyadenylation signal
PRE	Pumilio Response Element
PTC	Premature Stop Codon
PUM1/2	Pumilio 1 and 2
qRT-PCR	quantitative Real-Time PCR
RNA	Ribonucleic acid
RNAi	RNA interference
SD	Standard Deviation

shRNA	short hairpin RNA
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
ssRNA	Single-stranded RNA
TRC	The RNAi Consortium
uORF	upstream Open Reading Frame
UTR	Untranslated region
WT	wild type/intact
yGFP	GFP codon optimized for yeast

CHAPTER 1

INTRODUCTION

1.1 Post-transcriptional gene regulation

Precise control of protein expression is essential for healthy development and homeostasis for all organisms. Every step in the protein biogenesis pathway is highly regulated, with mechanisms in place to regulate the rate at which the gene is transcribed into mRNA, the mRNAs processing - splicing, addition of 5′ cap and cleavage and polyadenylation - and export to the cytoplasm. Once in the cytoplasm, other mechanisms control the rate of mRNA decay, the subcellular localization and the rate at which the mRNA is translated into protein, and the rate at which the protein is degraded. Though transcription rates often explain the majority of the variation in steady-state protein levels [1], post-transcriptional gene regulation also plays a crucial role and allows for more complex and responsive regulatory strategies. For example, mRNAs with fast decay rates respond more quickly to stimuli and reach new steady-state levels faster [2, 3].

The information required for the correct localization, translation and decay of mRNAs is found within the mRNA itself, in so-called *cis*-regulatory sequence elements (*cis*-elements). These sequence motifs are bound by trans-acting factors such as RNA binding proteins (RBPs) and non-coding RNAs like microRNAs (miRNAs). The trans-factors then interact with ubiquitous macromolecular complexes that localize, translate or degrade the mRNA. Although there are examples of cis-elements that are located within the coding region of the mRNA, most cis-elements reside within the untranslated regions (5 and 3 UTRs) [4]. The 3′ UTR is particularly rich in regulatory information due to its size (average ~1300 bases in humans)[5]and relatively high number of conserved regions [6, 7]. The importance of the 3′ UTR is best illustrated by the consequences observed when genes

lose portions of their 3′ UTRs, for example through regulated alternative polyadenylation (APA). Highly proliferative cells show a global shift towards the use of proximal polyadenylation signals (PAS), resulting in shorter 3′ UTRs [8]. Importantly, the same effect has also been observed in cancer cells, and alternative isoform usage for a specific gene, IGF2BP1, has been demonstrated to drive oncogenic transformation [9].

It is clear that the precise expression of the majority of genes is dependent on information encoded within their UTRs, particularly within the 3′ UTR. Each UTR may contain multiple regulatory sequence motifs and bind multiple different *trans*-acting factors, forming so-called messenger-ribonucleoprotein complexes (mRNPs). The composition of the mRNP is dynamic, changing depending on the cellular context and the stage of the mRNAs life-cycle, and controls the fate of the mRNA [10]. Given the tendency for most regulatory sequences to be located in the 3′ UTR, the 3′ UTR can be considered the main switchboard from which post-transcriptional events are regulated, at least in mammals. Here, I will discuss 3′ UTR-mediated post-transcriptional gene regulation, with a particular focus on regulatory elements in mammals, their qualities and their interactions, and on how all the information within a 3′ UTR comes together to determine an mRNAs fate.

1.2 Quantitative and qualitative regulation

3′ UTR-mediated post-transcriptional gene regulation can be either quantitative or qualitative. While quantitative regulation changes the amount of protein produced from the transcript, qualitative regulation affects the location or function of the protein (which as a consequence may also alter protein levels). Because our understanding of quantitative regulation is more advanced, many of the examples discussed in this chapter will be quantitative. However, recent evidence suggests qualitative regulation may be more common than previously anticipated. It remains to be determined whether mammalian

3′ UTRs function predominantly as quantitative or qualitative determinants of gene function.

1.2.1 Quantitative regulation

Quantitative post-transcriptional gene regulation mainly affects translation and mRNA decay rates, controlling how many protein molecules each transcript produces during its lifetime. Both translation and decay can be regulated by trans-factors bound at the 3′ UTR. In cap-dependent translation, the circularization of the transcript through interactions between the polyA binding proteins (PABPs) at the polyA tail and the 5′ cap binding complex, which includes eIF4E, eIF4G and the helicase eIF4A, brings the 3′ UTR into close proximity to the 5′ end of the mRNA [11]. The circularization of the mRNA allows RBPs bound to 3′ UTR regulatory elements to influence translation initiation, which is usually the rate-limiting step in translation. For example, in humans the binding of the IFN-γ-activated inhibitor of translation (GAIT) complex to a structural element in the 3′ UTR of the copper-carrying glycoprotein ceruplasmin (Cp), leads to translational silencing of Cp [12]. In this case, a component of the GAIT complex, L13a, interacts with eIF4G of the translation initiation complex and blocks the assembly of the 48S ribosome subunit [13].

In general, mRNAs are protected from decay by their 5′ caps and their polyA tails. Decay, therefore, is usually initiated by complexes that shorten the polyA tail (deadenylases) or remove the 5′ cap (decapping enzymes), leading to decay by exonucleases, such as the exosome and XRN1. Deadenylases and decapping enzymes can be recruited to the mRNA through regulatory elements in the 3′ UTR. For example, hnRNPA2/B1 has been shown to bind specific regulatory sequences in many human 3′ UTRs and recruit the CCR4/NOT deadenylase complex, leading to mRNA decay [14].Increasing the rate of decay is also the primary mode by which microRNAs (miRNAs) regulate gene expres-

sion, though they also inhibit translation [15]. In complex with argonaute (Ago) proteins, miRNAs form the miRISC complex, whose binding to mRNA targets leads to translational repression and increased rates of mRNA decay. There are hundreds of miRNAs in mammals [16], and each miRNA can target hundreds or thousands of mRNAs [17]. In fact, a majority of mammalian mRNAs are conserved targets of miRNAs [18]. Through miRNAs and other 3UTR trans-factors that mediate quantitative regulation, quantitative post-transcriptional regulation affects a large majority of protein coding transcripts.

1.2.2 Qualitative regulation

The most well-studied example of qualitative 3′ UTR-mediated post-transcriptional regulation is undoubtedly regulation of mRNA localization. The localization and local translation of mRNAs is extremely important, particularly in development and in highly polarized cells such as neurons, where it helps establish cell polarity and allows quick responses to stimuli [19].Recently, it has become clear that many mRNAs that are localized. For example, thousands of mRNAs have specific localization patterns in *Drosophila* embryos [20] and in rat neurons [21, 22], as do hundreds of mRNAs in Xenopus tropicales embryos [23]. Beyond roles in establishing cell polarity, mRNA localization is also important for migrating cells such as fibroblasts and certain epithelial cells [24, 25]. It remains to be seen how widespread mRNA localization is in other less extremely polarized cells.

Though mRNA localization is certainly a type of qualitative regulation, as it controls the location of the resulting protein, many localization events also have quantitative consequences. For example, mRNAs whose localization is achieved by active transport are translationally silenced until they have reached their destination. For some mRNAs, even when they have been correctly localized, activation of translation requires a specific signaling event [26, 19]. For other mRNAs, localization is achieved by asymmetric degrada-

tion, whereby the transcript is protected from decay only within a specific region of the cell [27, 28]. Therefore, a crucial component of both of these localization mechanisms is quantitative.

Beyond their role in mediating RNA localization, very little is known about qualitative regulatory roles that 3′ UTRs may play. Recent evidence, however, hints that such events may be common. A study in the Mayr laboratory illustrated a novel role for 3′ UTRs in protein localization, one which is independent of mRNA localization. Here, the 3′ UTRs of five different mRNA transcripts act as scaffolds for a protein complex that is then recruited to the nascent polypeptides being translated from the mRNA. This diverts the newly synthesized proteins to the plasma membrane rather than the endoplasmic reticulum (ER), changing their function [29]. In this case, *cis*-elements within the 3′ UTR change the localization of the encoded protein by recruiting the necessary *trans*-factors to the site of translation. However, it is easy to imagine that such *trans*-factor recruitment from a 3′ UTR to the nascent polypeptide could be more widespread, with diverse functional consequences.

Recent proteomic studies provide further evidence indicating a possible widespread qualitative function of 3′ UTRs. These studies have identified many enzymes among mRNA bound RBPs (mRBPs), including kinases, E3 ubiquitin ligases, and metabolic enzymes [30, 31, 32, 33, 34, 35, 36]. As the methodology used in these studies does not distinguish functional from spurious binding events, many of these enzyme-binding events may be non-functional. However, given the large number of enzymes identified as novel RBPs, some are likely to have physiological consequences. In fact, some of the RNA-binding enzymes show RNA binding activity in both yeast and human, indicating that this is a conserved function of the enzymes [36]. A subset of RBP enzymes has been demonstrated to quantitatively regulate the mRNAs they bind. These enzymes are called moonlighting enzymes, as in addition to their well-defined enzymatic roles, which are

independent of RNA, they can also function as traditional mRBPs [37, 38]. For others mRBP enzymes, it is conceivable that the RNA regulates the enzyme's activity with proposed mechanisms including competition with the substrate for the enzymes active site and allosteric regulation [10, 38]. Though no such qualitative example involving mRNAs has been identified, examples exist of enzyme activation through binding to tRNA and virally derived dsRNA [39, 40, 10, 41, 38]. In addition, the assembly of glycolytic enzymes into a higher-order complex is dependent on RNA [42], indicating an RNA species functions as a scaffold to bring those enzymes together. These non-mRNA examples serve as a proof-of-principle that RNAs can play a role enzyme regulation.

As quantitative regulatory mechanisms have more easily quantifiable consequences, most regulatory sequences discovered to date fall within this category. Apart from mRNA localization, whose importance is quite clear, the extent to which other qualitative regulatory mechanisms contribute to the function of a typical 3′ UTR remains unclear. The discovery of 3′ UTRs functioning as scaffolds for the recruitment of protein complexes to the nascent protein, and the discovery of many enzymes with no clear connection to RNA processing among mRBPs, holds exciting promise for further discoveries of qualitative regulatory mechanisms. Though I believe the potential novel qualitative regulatory mechanisms are unlikely to overshadow the quantitative regulatory roles of 3′ UTRs, it is likely they will prove to be both diverse and wide-spread, with important physiological consequences.

1.3 Anatomy of a 3′ UTR *cis*-regulatory element

The identification and functional characterization of regulatory *cis*-elements in 3′ UTRs has been an active field of research for the past couple of decades. *Cis*-elements have been identified computationally based on conservation or overrepresentation of sequences

[43, 44, 45, 46, 6] or enrichment of motifs in subgroups of co-regulated mRNAs [47, 48], biochemically through *in vitro* [49, 50] or *in vivo* [51, 52, 53, 54, 55] binding affinities to specific *trans*-acting factors, or functionally through reporter assays [56, 57, 58]. Recent advances in high-throughput technologies have lead to the identification of hundreds of potential *cis*-elements. However, the functions of the majority of these elements remain to be characterized. Nevertheless, the enormous number of identified *cis*-elements allows us to start seeing general patterns to regulatory sequences recognized by *trans*-factors. Three primary parameters of *cis*-elements need to be considered: primary sequence, secondary structure, and post-transcriptional editing and modifications.

1.3.1 Primary sequence

Though there are *trans*-factors that bind mRNAs indiscriminately, recognizing either single-stranded RNA (ssRNA) or double-stranded RNA (dsRNA), and others that recognize mRNAs exclusively through secondary or tertiary structure, most have at least some preference for specific sequences. *Cis*-elements that are primarily recognized through primary sequence are usually short (6-12 nucleotdes (nts)) [43] and range between well-specified sequences to ones with multiple degenerate positions.

Perhaps the most well studied *cis*-regulatory elements, and amongst the most sequence-specific, are target sites for microRNAs (miRNAs). In complex with argonaute (Ago) proteins, miRNAs form the miRISC complex, whose binding to mRNA targets leads to translational repression and increased rates of mRNA decay. The miRISC is very sequence specific due to the fact that miRNAs bind their targets through base-pairing interactions, rather than depending on interactions between amino acids, or protein structures, and RNA bases. Canonical miRNA target sites can be predicted based on base-pairing to the seed region of the miRNA (nts 2-7), with either an A opposite position 1

and/or a match to position 8. Additional basepairing to positions 13-16 can also contribute to targeting [59]. Though non-canonical miRNA target sites have been identified and their binding verified [51, 60], another study found evidence indicating that those sites are generally not functional [61]. Taken together, these studies indicate that functional non-canonical sites certainly exist but that they are greatly outnumbered by canonical seed-type sites. Other highly sequence-specific motifs include Pumilio Response Elements (PREs), bound by proteins containing Pumilio-homology domains. An example in humans are PUM1 and PUM2, which bind to UGUANAUA motifs, leading to repressed (or in special cases enhanced) expression of their target [62].

Less specifically sequence-determined target sites are also quite common. An example are AU-rich elements (AREs), a family of regulatory elements that can be bound by a whole cadre of RPBs (so-called ARE-BPs). The strict definition of an ARE is an element that is rich in As and Us that contains at least one instance of an AUUUA motif. However, many ARE-BPs, though they do bind U-rich regions, do not require the presence of this pentamer. Some ARE-BPs destabilize mRNAs (TTP, BRF1, KSRP), others stabilize them (HuR, HuD, CUG-BP2, Hel-N1, Nucleolin, PAIP2), and some (AUF1) can either stabilize or destabilize depending on the target [63, 64]. Similar to AREs are GU-rich elements (GREs), which are overlapping clusters of GUUUG motifs, as opposed to AUUUA motifs for AREs. GREs are bound by CELF1, among others, and mediate decay [64]. Both AREs and GREs illustrate an important fact, true of many *cis*-regulatory elements, that a single site may be bound by many different *trans*-factors, leading them to have complex and often contradicting effects.

RBPs that recognize dsRNA (dsRBPs) have traditionally been thought to mostly recognize the shape and structure of the RNA molecule, recognizing the A-form helix and 2OH groups of the backbone [65]. Though A-form helices have narrow major grooves, the information in the more accessible minor groove is enough for dsRBPs to have some

sequence preferences. A good example of this is the A-to-I RNA modifying enzyme ADAR2, which can edit RNA in a non-specific or sequence specific manner. The structure of ADAR2 bound to an mRNA target in solution showed that the protein interacts with specific basepairs in the minor groove and that those interactions are important for high affinity binding and efficient editing [66]. Thus, many RBPs likely recognize elements based on both their structure and their sequence.

The examples given above are only a small subset of the hundreds of sequence-based regulatory elements that have been identified, but they illustrate some general principles of *cis*-element recognition. First, that *cis*-elements can range from having very specific sequence requirements to being more loosely determined. Second, that a single *cis*-element may be bound by many different *trans*-factors, with different consequences for gene expression. And third, that even regulatory elements that are primarily structure determined may also have specific sequence requirements.

1.3.2 Secondary structure

The extent to which 3′ UTR *cis*-element recognition is driven by secondary structure is still very much up for debate. Some elements, particularly those bound by dsRBPs, are clearly primarily or exclusively structure determined (e.g. many zipcodes - elements that determine mRNA localization - and binding sites for RBPs ADAR2 and Staufen). For others, the role of structure is less clear. Many RBPs strongly prefer to bind ssRNA and some even require an unstructured local context. However, some such proteins require or prefer that the ssRNA element be presented in the loop or bulge of a hairpin [67]. See figure 1.1 for examples of regulatory element structures (Fig. 1.1)

As examples of *cis*-elements whose binding is primarily structure-determined, I will describe two *cis*-elements, the beta-actin zipcode and Staufen binding sites. The beta-actin

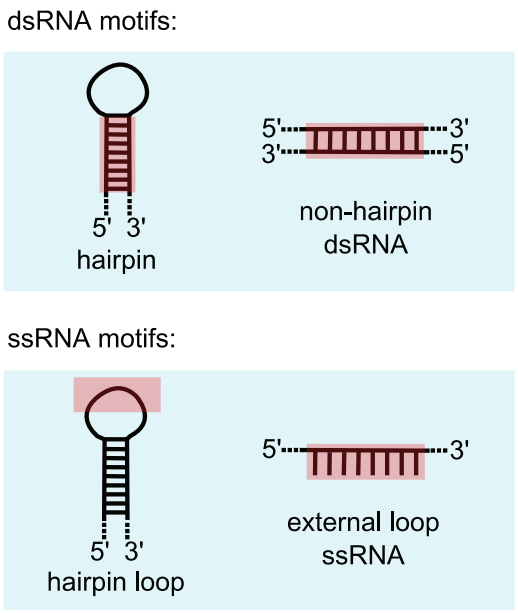


Figure 1.1: Examples of *cis*-element structures

zipcode forms a stem-loop with two internal bulges and a 16-nt terminal loop. Its binding partner, ZBP1, specifically recognizes this structure with the specific sequence playing a much smaller role. This was established by examining the effect of mutations that disrupt the sequence but maintain the structure and vice versa. Changing the structure disrupted ZBP1 binding but changing the sequence (while maintaining the structure) did not [68]. Another RBP that regulates mRNA localization and primarily recognizes structure is Staufen, though it also has reported roles in mRNA stability and translation [69]. Staufen recognizes dsRNA through its four dsRNA binding domains. The RNA duplexes recognized by Staufen are usually between 5 and 14 nts in length and have an asymmetry with purines on one arm and pyrimidines on the other [55]. Staufen binding sites also have another quality that is different from the beta-actin zipcode. While the beta-actin zipcode is defined by a specific stem-loop structure within a 3′ UTR, Staufen also binds dsRNA formed by long-range interactions or even intermolecular base pairing between target mRNAs or between a target mRNA and a lncRNA [70]. A recent study found a large number of long-range interactions among Staufen binding sites, mostly within 3′

UTRs but also between different RNA molecules. While it remains to be seen how many of these binding events are functional, evidence from Single Nucleotide Polymorphisms (SNPs) indicates that disrupting these sites has deleterious consequences that are selected against [55].

In addition to RBPs that specifically recognize RNA secondary structure, there are ssRNA binding proteins that have higher binding affinities to motifs that are present in the loop or bulge of a hairpin [67]. CPEB3 and CPEB4 (CPEB3&4) are a specific example. CPEB3&4 are related to CPEB, a protein that mediates cytoplasmic polyadenylation of target transcripts to relieve translational repression [71], though they have slightly different functions and their binding affinities are quite different. While CPEB binds its targets in a strictly sequence dependent manner, CPEB3 and 4 both require that the U-rich recognition sequence be presented in a hairpin bulge [72].

Though specific examples of ssRBPs with structure preferences exist, the extent to which this applies to many RBPs in not clear. However, a recent study by Ray et al. started to address this question. They did a comprehensive *in vitro* RNA binding experiment (RNAcompete) for 207 different RBPs and queried structure preference for dozens of them. None of those absolutely required a specific structure, but seven of them showed a preference for hairpin loops and twenty-two had a bias against them [73]. Additionally, a smaller study by Li et al. determined to what extent structural context (location in hairpin loops vs. unstructured external loops) improved target site prediction. Of the nine proteins queried, five were more strongly predicted within external loops and only one favored hairpin loops [67]. Together, these studies indicate that requirements for structured context (e.g. within hairpins) is rare. Conversely, they argue that an unstructured context is necessary for the recognition of many *cis*-elements.

1.3.3 RNA editing and modification

A fairly recent development in the study of 3′ UTRs is the discovery of the importance of RNA editing and modification for the binding of certain mRBPs. While editing events change one base into another, RNA modifications produce a base that is not identical to a nucleotide produced separately by the cell. In addition, RNA modifications are generally more dynamic, as there are specific enzymes that can remove them.

By far the most common editing event is A-to-I editing of dsRNA mediated by adenosine deaminase acting on RNA (ADAR) [74, 75]. Though introns and intergenic regions are the most enriched in A-to-I editing events, 3′ UTRs are also quite commonly edited, comprising 20% of the cytosolic editome in polyA+ data from human ENCODE cell lines [74]. As A and I have different base-pairing properties (A pairs with U, and I pairs primarily with C, though it can also pair with A and U), these editing events can change *trans*-factor binding if they fall within *cis*-elements. For example, Gu et al. found that editing events are enriched within miRNA target sites in several mouse tissues [76] and Chen found that 3′ UTR edited sites were enriched in events that disrupt, create, or switch miRNA targeting, when compared to random sites [74]. Specific instances of miRNA disruption by A-to-I editing have been confirmed *in vivo* [77, 78]. Though no instances of A-to-I editing disrupting RBP binding have been reported (to my knowledge), it is easy to imagine that such events must take place, given the high rate of editing and the high number of RBP binding sites. Another way in which A-to-I editing in the 3′ UTR influences *cis*-element recognition is by allowing long dsRNA motifs to avoid triggering apoptosis [79]. Therefore, these elements could not exist without editing. A second, less common, RNA editing event is cytidine deamination by APOBEC proteins to create a C-to-U alteration [80, 81, 82]. C-to-U editing events are also enriched in 3′ UTRs, and particularly within AU-rich segments [83].

Modified bases commonly found in mRNAs include pseudouridine (Ψ) [84, 85], 5-methylcytosine (m5C) [86, 87] and N6-methyladenosine (m6A) [88, 89]. By far the most common modification in mRNAs is m6A, so I will focus on that here. M6A is put in place by the METTL3/14 complex and removed by FTO or ALKBH5 [90, 91, 92]. On average, each mRNA contains 3-5 m6As [91] and, like the editing events described above, this modification is enriched in 3′ UTRs [88, 89, 93]. Interestingly, a connection has been made between m6A modification and miRNA targeting. Not only are m6As enriched at miRNA target sites, one study found that changes in the levels of miRNAs lead to changes in m6A modification levels at the target sites of the affected miRNAs [89, 94].

The first proteins found to bind specifically to m6A containing sequences were YTHDF1-3 [88]. YTHDF1 promotes translation and YTHDF2 regulates mRNA stability, and their binding sites are primarily in the 3′ UTR [95, 96]. Another RBP that has been shown to bind m6A containing motifs is hnRNPA2B1. In this study, the effect of hnRNPA2B1 at modified sites was to regulate alternative splicing and miRNA biogenesis [97]. However, recent work from the Grimson lab identified binding sites for hnRNPA2B1 in the 3′ UTR that cause mRNA decay [14]. It remains to be determined if the identified hnRNPA1B1 binding sites in the 3′ UTR are edited.

Overall, though most mRNA-binding RBPs primarily recognize primary sequence, it is clear that both secondary structure and RNA editing or modifications are often extremely important and should not be ignored.

1.4 The effect of local context on *cis*-elements

Beyond the sequence, secondary structure, and modifications present in the regulatory elements themselves, one must also consider the immediate surrounding sequence con-

text. Even for miRNAs, whose binding is based on predictable base-pairing interactions, sequence information alone cannot perfectly predict which target sites are functional [59]. Two major sequence-context parameters have the largest effects: 1) Accessibility and 2) the nearby presence of other sequence elements.

1.4.1 Accessibility

As mentioned above, an unstructured context is important for binding of many *trans*-regulatory factors. For many ssRBPs, it is necessary that their cognate motif be accessible in a single stranded conformation for recognition and/or binding to occur. For miRNAs, the role of accessibility in target site choice has been clearly established [98, 99]. In addition, binding site prediction models have increased accuracy when they incorporate either structural information or AU-richness in the surrounding sequence, which generally indicates a lack of structure [100, 59]. More recently, local structure predictions have also been demonstrated to improve binding site predictions for many ssRBPs [67, 101, 102]. The important implication of these findings is that binding to these elements can be regulated by changing the local secondary structure. One particularly interesting example of such a regulatory mechanism is mediated by m6A modification. Because an m6A modified base has decreased base-pairing affinity [103], modifications near binding sites for the RBP HNRNPC increase their accessibility and allow binding to occur [104]. Finally, local structure can also be influenced by *trans*-factor binding to neighboring sites, as discussed below.

1.4.2 Nearby regulatory sequences

The second potential local influencer is the presence of nearby *cis*-elements whose binding by *trans*-factors can change the binding affinity to the site of interest. For instance, two closely spaced miRNAs can function synergistically, enhancing their regulatory impact beyond expectation for independent sites [59, 105, 106]. Furthermore, there have been multiple reports of RBPs influencing miRNA repression [107, 108, 109], and of RBPs interacting with one another [108]. These interactions primarily occur through the following mechanisms: 1) cooperative binding, 2) modulating accessibility, and 3) competition/steric hindrance (Fig. 1.2).

In a cooperative binding interaction, interactions (hetero and homo) between *trans*-factors facilitate binding to adjacent regulatory *cis*-elements. An informative example of this is the interaction between FMRP and MOV10 on their shared mRNA targets in the mouse brain [54]. FMRP and MOV10 are both RBPs that had previously been shown to have connections with miRNA-mediated translational repression, though the mechanism was unclear. Kenny et al. demonstrated a physical interaction between FMRP and MOV10 that is only partially RNA dependent, indicating that protein-protein interaction is required. They also showed that MOV10 binding to shared targets is dependent on FMRP, as binding is dramatically reduced in FMRP knockouts. Interestingly, cooperation with FMRP changes the effect of MOV10 binding on miRNA mediated repression, going from enhancing repression on its individual targets to impeding repression on targets shared with FMRP [54]. The cooperative binding of RBM38 and HuR on p21 is another, quite similar, example [110].

Another way for synergism between *cis*-elements to occur is for the binding of one *trans*-factor to change the local RNA secondary structure to increase accessibility to a second nearby site. Two studies have shown that the human PUF family proteins Pumilio 1

and 2 (PUM1/2) can enhance miRNA repression using this mechanism [111, 112]. Kedde et al. studied the tumor suppressor p27, whose repression is essential for cell cycle entry. In proliferating cells, p27 repression is mediated by the miRNAs miR-221 and miR-222. In quiescent cells, the target sites for miR-221 and miR-222 are paired to a Pumilio Response Element (PRE). The pairing creates a hairpin, which reduces the accessibility of the miRNA target sites and inhibits the repression of p27. Upon cell cycle activation through growth factor stimulation, PUM1/2 binds to the mRNA causing a structure switch that releases the miRNA target sites from the hairpin, allowing for p27 repression and cell cycle entry [111]. A similar structure switch mechanism was also proposed between PUM1/2 and multiple miRNAs targeting the E2F3 oncogene, and disruption of this interaction was noted in multiple cancer cell lines [112]. This is also the mechanism by which PTB facilitates miRNA repression of GNPDA1 and most likely how MOV10 enhances miRNA mediated repression, as it has helicase activity that allows it to unwind secondary structures [113, 54].

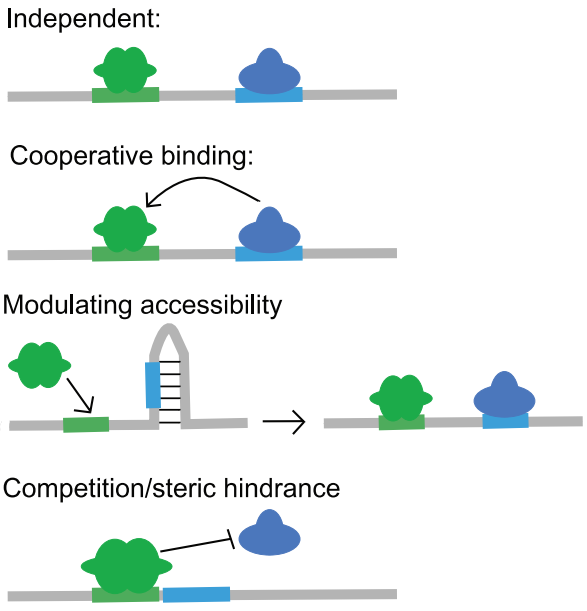


Figure 1.2: Main mechanisms of *cis*-element interactions

For other demonstrated cooperative events, the exact mechanism of cooperation has

not been elucidated, though they likely utilize one of the two mechanism described above. This includes synergistic miRNA target sites [59, 105, 106] and the HuR mediated recruitment of *let-7*-bound Ago proteins to the c-MYC 3′ UTR [114].

Antagonistic interactions between *cis*-elements usually involve competitive binding to a single site, or mutually exclusive binding to two very closely spaced sites. AREs are a major site of this kind of competition, with different ARE-BPs competing for access to same the ARE. Often it is a competition between the stabilizing HuR and one of the many destabilizing ARE-BPs, such as AUF1, TTP or CELF [115, 116, 117]. Competition of this sort has also been demonstrated between miRNA target sites and binding sites for RBPs. A good example is the RBP DND1, which has been shown to inhibit binding of miRNAs to nearby target sites in several mRNAs [118]. Other examples include the competition between hnRNP L and miRNAs at a CA-rich element (CARE) in the 3′ UTR of VEGFA [119] and the competition between hnRNPA2B1 and HuR on the CDK5R1 3′ UTR [120].

For other antagonistic interactions the mechanism is not as clear. HuR, for example, has been found to relieve miRNA-mediated repression in several contexts [121, 122, 123, 124], only some of which had HuR binding to AREs in close proximity to the miRNA target sites. Kundu et al. proposed that the ability of HuR to oligomerize on the mRNA allows it to displace bound miRNAs even at a distance [123].

Though these examples of *cis*-element interactions are certainly compelling, it remains unclear to what extent interactions are a general paradigm for 3′ UTR regulatory elements. To attempt to answer this, we need a fuller picture of 3′ UTRs as a whole.

1.5 Towards a picture of a 3′ UTR

To understand post-transcriptional gene-regulation, we need to look beyond individual regulatory elements and *trans*-factors and examine the 3′ UTR as a whole. To get a clear picture of how an entire 3′ UTR functions, we need to answer three questions: 1) How densely is it packed with regulatory elements, 2) how independently do those regulatory elements function, as a rule, and 3) how structured is the 3′ UTR? Only with answers to these questions can we comprehend the role of post-transcriptional regulation mediated by the 3′ UTR in gene regulatory networks.

1.5.1 Density of regulatory information

There are several different ways to begin to determine how dense with regulatory information a typical 3′ UTR is. One can consider the proportion of the possible sequence space that has regulatory potential, the proportion of a 3′ UTR that is conserved, *trans*-factors and how densely a 3'UTR is bound by them, and one can directly measure the regulatory effect of each section using a 3′ UTR bashing approach.

One way to determine the proportion of sequence motifs that have regulatory function is to examine the data from high-throughput functional screens. Several labs have recently done reporter screens using fluorescence activated cell sorting (FACS), coupled with high-throughput sequencing, to measure the regulatory potential of a large number of sequences [56, 57, 58]. Only one of these screens used completely random sequences, making it easier to determine the ratio of functional to non-functional motifs, so I will focus on that study here. Wissink et al. examined ~8000 random 8nt motifs (8mers) in a large pilot screen and identified ~1,100 8mers as potential regulatory motifs. This indicates that about 14% of random 8nt sequences are likely functional under their exper-

imental conditions [58]. The false positive rate in this experiment was quite high and it is difficult to establish the false negative rate for such experiments. Even so, this results serves as a good starting point.

Using a conservation-based approach, other groups reached very similar numbers [6, 7]. Conservation-based approaches rely on the fact that functional sequences are likely to be under purifying selection, making their mutation rates lower than the surrounding sequence. The groups used two different conservation metrics. One group used Hidden Markov Models to determine which bases are likely to be within a *cis*-regulatory element (PhastCons) [6]. The other group calculated the likelihood that each base is preferentially conserved (PhyloP) [7]. PhastCons estimated that 18% of bases in 3′ UTRs are members of a regulatory element and PhyloP estimated that 13% of bases are preferentially conserved. Importantly, these numbers exclude recently evolved functional motifs, which would not show preferential conservation.

Not all studies agree with the estimates made above. A recent study determined which parts of the mRNA are bound by proteins by UV crosslinking proteins to 4-thiouridine-labeled RNAs, isolating and fragmenting protein-bound mRNAs, and sequencing them. They then looked for the diagnostic U-to-C mutation that arise at crosslinking sites and found that about a third of Us had been cross-linked [31]. Given that 3′ UTRs are generally U-rich, this indicates more extensive binding to 3′ UTRs than hinted at by the analyses described above. However, not all protein binding events captured by such crosslinking experiments are functional [61], which may contribute to the difference in estimates.

The most direct approach to estimating the density of regulatory information is to directly measure the regulatory effect of sequences across a 3′ UTR at a high resolution. Very few 3′ UTRs have been studied to a high enough resolution. Zhao et al. generated an almost single nt resolution regulatory information map of the 695nt CXCL2 3′ UTR by measuring, in high-throughput, the effect of SNPs, generated by error-prone PCR, on

steady state mRNA levels. They found 5-9 regulatory regions, depending on the threshold used, translating to 7-13 sites per kb [57]. In work presented in this thesis, I analyzed the mouse *Hmga2* 3′ UTR and the human Pim1 3′ UTR to a 50 and 100nt resolution, respectively, by measuring the regulatory effect of isolated 3′ UTR fragments on reporter protein expression. For both 3′ UTRs, about 60% of the fragments assayed displayed regulatory effects, translating to at least 11.8 and 6.2 sites per kb, for *Hmga2* and Pim1 respectively [125]. Finally, though the Tavasoie laboratory did not systematically cover an entire 3′ UTR, their study of conserved segments of 3′ UTRs gave information that can be used to determine density of regulatory sequences. Using a high-throughput screen to measure a large number of 34-nt long conserved pieces of real 3′ UTRs, they identified 2050 of 16,332 3′ UTR pieces as having potential regulatory function, translating to at least 3.6 sites/kb [56]. The consistency between these experiments, with estimates well within an order of magnitude for different 3′ UTRs, is very encouraging.

To summarize, according to our current knowledge, at least 13-18% of 3′ UTR sequences are likely functional and 3′ UTRs contain between 3 and 13 *cis*-elements per kb. The identity and function of a large fraction of these sites remains to be determined.

1.5.2 The extent of *cis*-element interactions

There are many reported instances of 3′ UTR *cis*-elements that exhibit consequential interactions with other 3′ UTR *cis*-elements. A key question in 3′ UTR biology relates to the frequency of such interactions: are they a general paradigm for post-transcriptional gene regulation, or do they simply represent interesting exceptions to the typical function of 3′ UTRs? To start to answer this question, several computational studies looked for, and found, an enrichment of miRNA target sites in the vicinity of PREs [126, 127] and also select other RBP binding sites [102, 128, 129], possibly suggesting wide-spread inter-

actions between the miRNA silencing machinery and RBPs bound nearby. Of particular interest is the finding that PRE motifs are highly enriched near functionally validated miRNA target sites that were predicted to be structurally inaccessible [130]. Remarkably, miRNAs that possess target sites enriched in proximity to PREs tend to have target sites that are reverse complementary to the PRE, and can therefore form hairpins [131]. These findings strongly suggest that the PUM1/2-induced structure-switching mechanism described above for p27 is a general mechanism by which PUM1/2 regulate miRNA-mediated repression.

Not all RBPs queried in the studies discussed above displayed enrichment patterns indicative of interactions [102], demonstrating that some RBPs function mostly independently. In addition, miRNA target sites appear to retain consistent levels of efficacy in different cellular environments [132]. As different cellular environments imply differences in RBP expression levels, the consistency of miRNA targeting suggests that RBP-mediated modulation of miRNA targeting is unlikely to be extensive. Work presented in this thesis further demonstrates that interacting motifs are unlikely to be in the majority, as an intensive search for interactions across two separate conserved 3′ UTRs showed almost no evidence of interactions [125]. Overall, it seems that interacting regulatory motifs are fairly common, but not the general paradigm for post-transcriptional gene regulation.

1.5.3 Overall 3′ UTR structure

3′ UTR structures can both serve as binding sites for dsRBPs and affect the accessibility of *cis*-elements to their ssRBP *trans*-factors. It is therefore crucial to know to what extent 3′ UTR exist as linear or structured RNA. As 3′ UTRs are both long and densely occupied by proteins *in vivo*, determining their physiological structure is challenging. Recently, technical advances have allowed researchers to measure the global accessibility of RNA

nucleotides *in vivo* using DMS-seq or icSHAPE [133, 134]. In addition, the advent of hiCLIP allows researchers to identify paired RNA duplexes that are bound by specific RBPs [55], adding additional information. Though further study is needed to fully appreciate the role of 3′ UTR structure in gene regulation, these technical advances have helped the field make huge strides.

In vitro, most RNAs contain many basepairing interactions and form complex structures. However, biochemical experiments *in vivo*, both in yeast and in mouse ES cells [133, 134], found that mRNAs were less structured than expected based on earlier *in vitro* work, particularly at the 3′ UTR. In fact, Rouskin et al., using DMS-seq, found evidence of active unfolding of mRNA structures in yeast, whereby ATP depletion lead to an increase in mRNA structures. Overall, they found that about 4% of the regions examined were very likely to be structured and about 29% of the mRNA regions they examined were completely indistinguishable from denatured RNA, indicating that a large proportion of the 3′ UTRs is mostly unfolded [133]. However, the Chang laboratory illustrated the importance of many structures by studying so-called riboSNitches, single nucleotide variations (SNVs) that change RNA secondary structure. They found that 15% of transcript SNVs were riboSNitches, of which 211 were associated with changes in gene expression and 22 were associated with human diseases or other phenotypes [135].

Though biochemical measurements of ribonucleobase accessibility are extremely helpful, they still rely on computational structure prediction to translate accessibility information into potential structures. As the paired RNA duplexes get further apart, they get harder to predict, making it very difficult to gauge how common long-range interactions are. To accomplish this, the paired RNA duplexes need to be directly identified. A recent study using hiCLIP did exactly this, identifying the RNA duplexes bound by Staufen1. They found that long-range interactions were not uncommon, with 57% of Staufen1-bound 3′ UTR duplexes having loops larger than 100 nts, and many duplexes

even spanning the entire 3′ UTR [55]. Though it is quite possible that Staufen1 specifically favors long-looped structures, these data suggest that long-range interactions may be much more common than previously thought.

In conclusion, it is clear that while 3′ UTRs are less structured than *in vitro* studies predicted, they do contain many basepaired nucleotides that form both smaller hairpins and long-range interactions. However, it is important to keep in mind that RNA structures are highly dynamic and that they can be controlled by more than thermodynamic stability.

1.6 Adding layers to the 3′ UTR picture

With a framework for understanding 3′ UTRs emerging, there are additional parameters to be considered. These include the observed biases in the positioning of elements within 3′ UTRs, the effect of alternative polyadenylation and cleavage (APA) on 3′ UTR-mediated regulation, the potential role of 3′ UTR length *per se* as a regulator of gene expression, and finally the role of the polyA-tail and its modifications. Studies indicate that each of these parameters impacts 3′ UTR function. However, it is far from clear how widespread and important such impacts are. In figure 1.3, I have sketched out a model 3′ UTR with many important parameters highlighted (Fig. 1.3).

1.6.1 Regulatory element position

The effect of *cis*-element positioning within a 3′ UTR has not been widely considered. A handful of specific sites have been demonstrated to function better when located within a specific region of the 3′ UTR. A prime example are miRNAs, whose functional target sites tend to reside near either edge, at least in long 3′ UTRs [136, 59]. In addition, a newly

discovered binding site for hnRNPA2/B1, which induces mRNA decay by recruiting the CCR4-NOT deadenylation complex, has an even stronger preference for the edges of long 3′ UTRs [14]. This position bias, at least towards the 5′ tail of the 3′ UTR may be connected with m6A modifications, as those are also preferentially located near the start of 3′ UTRs [88, 89, 93]. The reason for this bias in regulatory element location is still unclear, but may be due to a functional reliance on proximity to either the stop codon or to the polyA-tail. Conversely, it may hint at an additional layer of regulation involving APA, whereby sites close to the stop codon are constitutively included and sites close to the polyA-tail are easily excluded.

A final example of position bias involves nonsense-mediated decay (NMD), a quality control and regulatory mechanism which targets and degrades, among others, transcripts with particularly long 3′ UTRs [137]. It has recently been discovered that certain motifs, a good example being PTBP1 binding motifs, can protect long 3′ UTR transcripts from NMD when located within about 200 nts of the stop codon but not when located elsewhere in the 3′ UTR [138, 139]. Again, the mechanism underlying this position-sensitivity is unclear, but as NMD depends on interactions between the NMD complex and the terminating ribosome [140], it is conceivable that PTBP1 interferes with this interaction. It will be critical to determine if other regulatory elements also display location biases, and whether such biases are the rule for 3′ UTR *trans*-factors.

1.6.2 Alternative polyadenylation and cleavage

Many transcripts can express alternate versions of their 3′ UTRs due to Alternative Polyadenylation and Cleavage (APA). APA can have important consequences for gene expression, as it changes the UTR's length and the complement of regulatory elements within it. According to recent estimates, about 70-75% of human genes can be subjected to

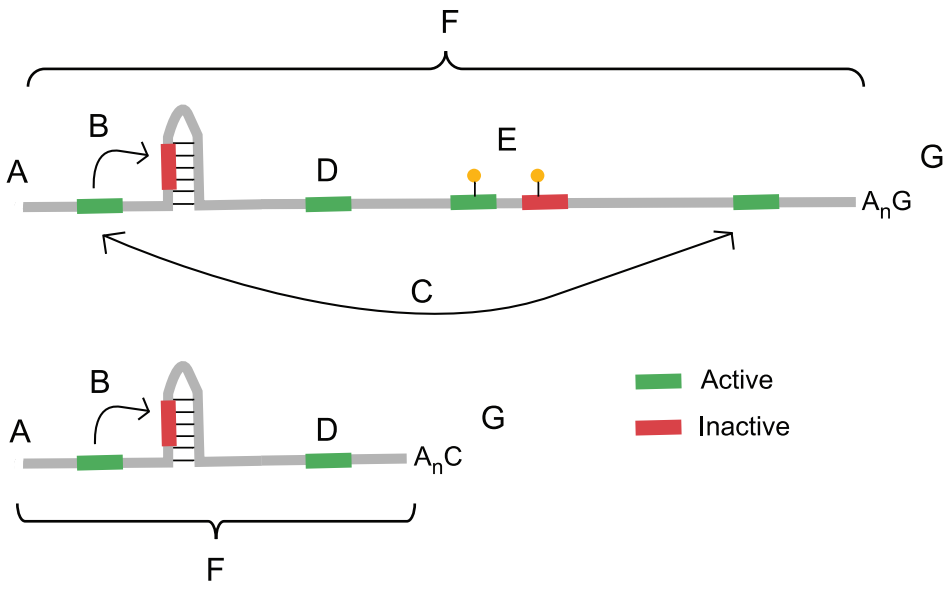


Figure 1.3: Model 3′ UTR with major *cis*-element and other 3′ UTR parameters highlighted (A) This model 3′ UTR has two isoforms formed through APA. (B) Short range interaction between two neighboring elements. Binding of *trans*-factor to upstream element increases the accessibility of the downstream element to its *trans*-factor by releasing hairpin structure. (C) Long-range interaction crossing the length of the 3′ UTR. (D) Effect of position within the 3′ UTR on *cis*-element effectiveness. *Cis*-element is inactive in long isoform, as it is located within the center. In the short isoform the *cis*-element is near the 3′ end and is active. (E) *Cis*-element editing or modification. Two sites are edited/modified in this 3′ UTR. One can only be bound by its *trans*-factor when edited/modified, while the modification of the second site inhibits binding by its *trans*-factor. (F) The effect of 3′ UTR length. The two isoforms of this 3′ UTR are very different in length. For the long isoform, its length may promote decay. (G) The effect of polyA-tail length and modification. The polyA tail can be shortened or lengthened with potential consequences for transcript stability. Addition of non-A nucleotides downstream of the polyA tail may protect it from deadenylation or promote transcript decay.

APA in one or more tissues or cell states [141, 142]. However, this estimate may overstate the frequency of alternate 3′ UTRs, as the underlying data used to inform the estimates (3′ end sequencing) had a relatively high rate of false positives, caused by recognition of internal A-stretches as polyA-tails [143]. It is nevertheless clear, that APA is extremely wide-spread and is, therefore, an important consideration when examining the role of 3′ UTRs on gene expression.

1.6.3 A role for 3′ UTR length

The extent to which 3′ UTR length contributes to gene expression regulation, independently of any regulatory sequences, has long been unclear. Nevertheless, the role of 3′ UTR length in NMD target selection has been clearly established. Apart from targeting transcripts with premature stop codons or upstream open reading frames (uORFs), NMD also targets mRNAs with particularly long 3′ UTRs, independently of premature termination [144, 145, 146, 137]. Hogg and Goff proposed that non-specific binding of UPF1, a key component of the NMD complex, measures the physical distance between the stop codon and the polyA-tail [147]. The exact length at which an endogenous 3′ UTR triggers the NMD pathway has been difficult to determine exactly, as the presence of certain *cis*-regulatory elements, including RNA structures that bring the polyA-tail closer to the stop codon, allows a subset of long transcripts to escape [139, 148]. For artificial 3′ UTRs, however, NMD is triggered at lengths as short as 420nt which is shorter than ~75% of human 3′ UTRs [148]. While this result indicates that the majority of human 3′ UTRs could be targeted by NMD, only 5-10% of the transcriptome is commonly estimated to be regulated by NMD [139]. It seems, therefore, that a large number of human 3′ UTRs evade NMD through known or unknown mechanisms.

Beyond its specific role in NMD, additional evidence exists that 3′ UTR length influences gene expression. The Bartel laboratory, while investigating the global contribution of APA to mRNA translation and decay rates, found that 3′ UTR length was negatively correlated with mRNA stability, explaining 9% of the variation they saw in half-life measurements in mouse fibroblasts [143]. However, an older study in another cell line did not find any effect of 3′ UTR length on mRNA stability [149]. This discrepancy may be due to a difference in techniques or it may be due to differences in the strength of the effect of 3′ UTR length in different cell types. A recent study in *S. cerevisiae* demonstrated a negative correlation between 3′ UTR length and polyA-tail length [150]. Though this correlation

did not translate to a corresponding difference in stability, a possible role in translation cannot be excluded as longer polyA-tails have been demonstrated to promote translation in yeast [151, 152]. Finally, in additional work presented in this thesis, we have discovered a potent and NMD independent negative correlation between the length of artificial 3′ UTRs and reporter expression. This effect was somewhat weaker when the 3′ UTR mimics were replaced with endogenous 3′ UTR sequences, indicating that endogenous 3′ UTRs contain features that allow them to escape the effect. Nevertheless, the identified decay pathway may very well play important roles in gene regulation in specific circumstances.

Overall, it seems clear that 3′ UTR length can have a role in determining the level of gene expression. However, it remains to be determined how commonly this is true, and to what degree the effect is dependent on NMD, as opposed to other decay pathways.

1.6.4 The polyA-tail and its modifications

Almost all mRNAs are polyadenylated at their 3′ end. A polyA-tail of a certain length is necessary for transportation of the mature mRNA into the cytoplasm, for efficient mRNA translation, and to avoid decay by 3′-to-5′ exonucleases. Several enzymatic complexes either shorten or lengthen polyA-tails, so polyA-tail length is dynamically regulated in the cytoplasm, though cytoplasmic shortening is the dominant paradigm in most situations [153]. Beyond the minimum length requirement, there are also several studies that indicate that the length of the polyA-tail is an indicator of either translation efficiency or the rate of mRNA decay. For example, polyA-tail length of hundreds of mouse liver mRNAs change along with the circadian rhythm and these correlate with circadian protein expression [154]. In addition, as mentioned above, the length of the polyA-tail has been connected to translation efficiency in yeast [151, 152]. Early zebrafish embryos also

display such a relationship between polyA-tail length and translation rates, but the connection is lost at later stages [155]. In mammalian cell lines, however, polyA-tail length correlates with mRNA half-life and not with translation efficiency [156, 155]. It is clear that more study is needed to determine the exact role of polyA-tail length and how and why this role is different in different cellular contexts.

The Kim lab discovered another potential regulatory mechanism involving the polyA-tail. They determined that uridylation, and less commonly guanylation, at the ends of polyA-tails is widespread, usually at a fairly low frequency for each transcript. Uridylation was preferentially located at short polyA tails and correlated weakly with lower half-life, and guanylation was preferentially located at long polyA tails and correlated weakly with longer half-life. Therefore, Chang et al. speculated that uridylation is involved in mRNA decay, a speculation that was confirmed in a later study, and that guanylation stabilizes mRNAs by inhibiting deadenylation [156, 157].

In summary, when considering the role of a 3′ UTR in gene regulation, one must take into account not just the complement of *cis*-regulatory sequences, but also their location within the 3′ UTR. One must also determine whether the transcript is subjected to APA and, if so, which isoform is dominant in the context of interest. In addition, the overall length of the UTR may also be an important consideration, especially if the 3′ UTR of interest may be a target of NMD. Finally, the length of the polyA-tail and how it is regulated and modified may be an important consideration.

1.7 Conclusion

Regulation of gene expression is extremely complex and functions at every level of the gene expression pathway. Post-transcriptional events are largely encoded within the

mRNAs 3′ UTR, which serves as the central switchboard that integrates the many signals it receives to produce the desired final concentration (or other qualitative outcomes) of the encoded protein. Another apt metaphor for 3′ UTRs is that they are a sentence, with each regulatory element forming a word. Though a word can have meaning on its own, the meaning of a word can change depending on the context. This context includes the words surrounding it, their position within the sentence, or the larger context within which the word is written (in this case the cellular context, such as the availability of *trans*-factors).

It has been the overarching goal of the work presented in this thesis to better understand how 3′ UTRs function as a whole, to try to move towards a sentence-level understanding of post-transcriptional gene regulation. Towards that goal I have worked on two major projects: The first was the complete characterization of the (quantitative) regulatory sequences within a single 3′ UTR, that of *Hmga2*, identifying every region with regulatory potential and ascertaining the extent to which regulatory elements interact with one another (Chapter 2). I found that the *Hmga2* 3′ UTR had many regulatory regions and that they generally functioned independently of each other. The second project was a study of the role of 3′ UTR length, independently of regulatory sequences, in the control of gene expression (Chapter 3). Using 3′ UTR mimics, random sequences with similar nucleotide composition to endogenous UTRs, I discovered a very strong effect of UTR length on mRNA stability, an effect which is much muted for endogenous 3′ UTRs. Importantly, though the mechanism remains mostly unclear, the effect is almost completely NMD-independent. Overall, I the results presented herein have brought us a step closer to an integrated understanding of 3′ UTR-mediated post-transcriptional gene regulation (Chapter 4).

CHAPTER 2

SYSTEMATIC ANALYSIS OF THE *HMGA2* 3′ UTR IDENTIFIES MANY INDEPENDENT REGULATORY SEQUENCES AND A NOVEL INTERACTION BETWEEN DISTAL SITES.

[1] [2]

2.1 Abstract

The 3′ untranslated regions (3′ UTRs) of mRNAs regulate transcripts by serving as binding sites for regulatory factors, including microRNAs and RNA binding proteins. Binding of such *trans*-acting factors can control the rates of mRNA translation, decay, and other aspects of mRNA biology. To better understand the role of 3′ UTRs in gene regulation, we performed a detailed analysis of a model mammalian 3′ UTR, that of *Hmga2*, with the principal goals of identifying the complete set of regulatory elements within a single 3′ UTR, and determining the extent to which elements interact with and affect one another. *Hmga2* is an oncogene whose overexpression in cancers often stems from mutations that remove 3′-UTR regulatory sequences. We used reporter assays in cultured cells to generate maps of *cis*-regulatory information across the *Hmga2* 3′ UTR at different resolutions, ranging from 50 to 400 nt. We found many previously unidentified regulatory sites, a large number of which were up-regulating. Importantly, the overall location and impact of regulatory sites was conserved between different species (mouse, human, and

[1] This chapter was adapted from a manuscript originally published in RNA and is reprinted with permission (Kristjánsdóttir K, Fogarty EA, Grimson, A (2015) Systematic analysis of the *Hmga2* 3′ UTR identifies many independent regulatory sequences and a novel interaction between distal sites, *RNA*, 21(7):1346-60)

[2] Though the experiments were designed by the author, Katla Kristjánsdóttir, many experiments (Fig. 2.1A,E,G,I,J; Fig. 2.2A,E; Fig. 2.3D, Fig. 2.4; Fig. 2.7 A,B,C) were executed in collaboration with Elizabeth Fogarty. Cloning and testing of constructs in Fig. 1.F was done by Andrew Grimson. Original primer design to make constructs in figure 1A was done by Raeanna Wilson.

chicken). By systematically comparing the regulatory impact of 3′-UTR segments of different sizes we were able to determine that the majority of regulatory sequences function independently; only a very small number of segments showed evidence of any interactions. However, we discovered a novel interaction whereby terminal 3′-UTR sequences induced internal up-regulating elements to convert to repressive elements. By fully characterizing one 3′ UTR, we hope to better understand the principles of 3′-UTR-mediated gene regulation.

2.2 Introduction

Precisely controlling the amount of protein made from each gene is a fundamental cellular process; while much of this regulation derives from transcriptional control, it is increasingly clear that regulation acting upon the mature transcript also plays a crucial role. The sequences underpinning post-transcriptional regulation are most often located within the transcripts 5′ and 3′ untranslated regions (UTRs). These *cis*-regulatory sequence elements within UTRs serve as binding sites for *trans*-factors such as RNA binding proteins (RBPs) and noncoding RNAs (for review see [158]). One relatively well-understood paradigm for post-transcriptional control is that of microRNAs (miRNAs), a class of ~22-nt small RNAs, which act in concert with the proteins of the RNA-induced silencing complex as *trans*-regulators of mRNAs. The effective binding sites for miRNAs are most often located within the 3′ UTR, and recruitment of a miRNA to a transcript results in transcript destabilization and translational repression (for review see [159]). While miRNAs are, perhaps, the most prevalent example of mammalian post-transcriptional regulation, a wide variety of other mechanisms exist. Hundreds of RBPs have been identified based on the presence of predicted RNA binding domains [160], but only a modest subset has been studied. Most post-transcriptional regulatory RBPs that have been studied have binding

sites within 3′ UTRs (e.g., [51, 52]). The 3′ UTRs of mammalian genomes, in particular, are typically both well conserved (compared with other regions of the genome; [6, 43]) and extensive in length (averaging 1.2 kb, versus 1.5 kb for coding sequences). Moreover, recent studies have indicated that 3′ UTRs are densely bound by proteins [31], many of which are likely to have regulatory roles. It seems the 3′ UTR serves as a switchboard that combines complex inputs leading to proper post-transcriptional regulation, a similar role to that of enhancers in transcriptional regulation.

While many, and likely most, 3′-UTR regulatory elements are uncharacterized, an increasingly large number have been identified. The majority of elements are between ~6 and 12-nt long and are recognized by *trans*-factors by virtue of their primary sequence [43]. Longer elements have also been described, whose identity and function derives from both structure and sequence [48]. Both site-types tend to be degenerate and therefore difficult to predict from examining the primary 3′-UTR sequence. However, some informative sequence motifs corresponding to *cis*-regulatory elements have been characterized genome-wide, including miRNA target sites, Pumilio Response elements, and binding sites for a small number of other regulatory proteins [161, 18]. It is, nevertheless, far from clear how many *cis*-regulatory elements exist in a typical 3′ UTR. Few 3′ UTRs have been systematically examined with the goal of identifying all of their regulatory sequences. However, studies that have begun to address this question suggest that 3′ UTRs might contain a large number of regulatory sequences; for example, mutational analysis of the *Caenorhabditis elegans cog-1* 3′ UTR shows that many parts of the 3′ UTR contribute to the post-transcriptional regulation of the cog-1 transcript [162]. Systematic studies of the regulatory landscape of 3′ UTRs to reveal the numbers and impacts of regulatory elements within them should enhance our understanding of the roles of 3′ UTRs in post-transcriptional biology.

The vast majority of investigations into 3′-UTR-mediated regulation has focused on

isolated, individual elements within a 3′ UTR and have not determined how collections of elements might interact with one another. However, there exist a handful of examples of 3′ UTRs containing combinations of *cis*-regulatory elements that function cooperatively: Closely spaced binding sites for miRNAs can function synergistically, enhancing their regulatory impact [59, 105, 106], and the binding of RBPs has been reported to both activate adjacent cryptic miRNA binding sites [114, 163, 111, 112] and mask otherwise functional sites [124, 118, 164, 123]. For example, two studies demonstrated that the binding of human Pumilio proteins within a 3′ UTR can induce a conformational change in the RNA structure. This structural change allows the silencing machinery access to miRNA target sites that were previously hidden within hairpins [111, 112] and thus enhances miRNA repression. Alternatively, the binding of RBM38 [164] or Dnd1 [118] within certain 3′ UTRs inhibits miRNA-mediated repression of mRNAs by inhibiting the binding of the silencing machinery, an example of an inhibitory interaction (or negative synergism). Importantly, many such interactions mediate major regulatory control of the underlying transcripts with important biological consequences.

These, and a handful of other examples, illustrate the potential importance of complex interactions between 3′-UTR regulatory elementsbut are such interactions rare or are they commonplace? At one extreme, if the impact of 3′ UTRs on expression can, typically, be derived from the independent contribution of individual elements, then a systematic approach to describe the role 3′ UTRs play in gene expression is relatively straightforward. Alternatively, if post-transcriptional regulation encoded within 3′ UTRs typically entails complex interactions between elements, then a genome-wide understanding of 3′-UTR-mediated regulation represents a nontrivial problem. In this study, we have systematically identified regulatory sequences within a single 3′ UTR, that of *Hmga2*, and developed a new approach to methodically determine the extent to which identified regulatory elements function independently or synergistically.

We selected the *Hmga2* transcript as a case study because it is strongly repressed post-transcriptionally through sites within the long (2.9 kb) and highly conserved 3′ UTR [165]. Normal expression of *Hmga2*, which encodes a nonhistone chromatin protein, is limited to embryonic tissues, and overexpression strongly correlates with poor prognoses for cancer patients. Moreover, *Hmga2* has been identified as a driver for metastasis [166]. One key component to the repression of *Hmga2* in normal nonembryonic tissues is the tumor suppressing miRNA *let-7*, which has seven target sites within the 3′ UTR. Loss of post-transcriptional regulation, and specifically loss of *let-7* targeting [167, 168], leads to overexpression of *Hmga2* and oncogenic transformation. Interestingly, the RNA binding protein IGF2BP3 has recently been shown to protect the *Hmga2* transcript from *let-7*-mediated repression in embryonic tissues and cancer cells [169]. While *let-7* repression and IGF2BP3 sequestration are clearly important, these alone do not explain the extensive conservation of the *Hmga2* 3′ UTR. Thus, it seems likely that additional regulatory sequences exist within what is already a relatively well-studied 3′ UTR.

A handful of 3′ UTRs have been examined systematically by truncation analysis with the goal of identifying important sequences. Most such studies were performed at low resolution (e.g., 400 nt), giving limited information about specific sequences [165, 170, 171, 172]. Two recent studies have taken a more comprehensive look at 3′ UTRs. Wirsing et al. (2011) first performed a low-resolution truncation analysis, and then performed higher resolution mapping of selected 3′-UTR fragments [173]. A high-resolution map of regulatory sequences within the CXCL2 3′ UTR has been generated by analyzing a large collection of point mutations within the 3′ UTR, though only measuring mRNA steady-state levels [57]. Although this study represents one of the most thorough investigations of a single 3′ UTR, focusing only on elements that control mRNA levels is not optimal, as elements that alter translation will be missed; moreover, elements robust to single point mutations would not have been detected, which may constitute a relatively large fraction of 3′ UTR elements. We have taken a different approach, map-

ping the *Hmga2* 3′ UTR to high-resolution (50 nt) using reporter assays that monitor total protein output. Importantly, we also measured the regulatory impact of larger overlapping fragments (100, 200, and 400 nt) and used comparisons between different sized 3′-UTR fragments to determine whether combinations of regulatory elements are interacting synergistically. We found a large number of previously unidentified regulatory segments, many of which confer up-regulation. Importantly, our data suggest that elements within the 3′ UTR largely function independently of one another. We did, however, observe an exception to this rule, whereby distal sequences within the 3′ UTR induce otherwise positive regulatory elements to function as repressive elements.

2.3 Results

2.3.1 High-resolution mapping reveals many discrete regulatory sequence elements within the *Hmga2* 3′ UTR

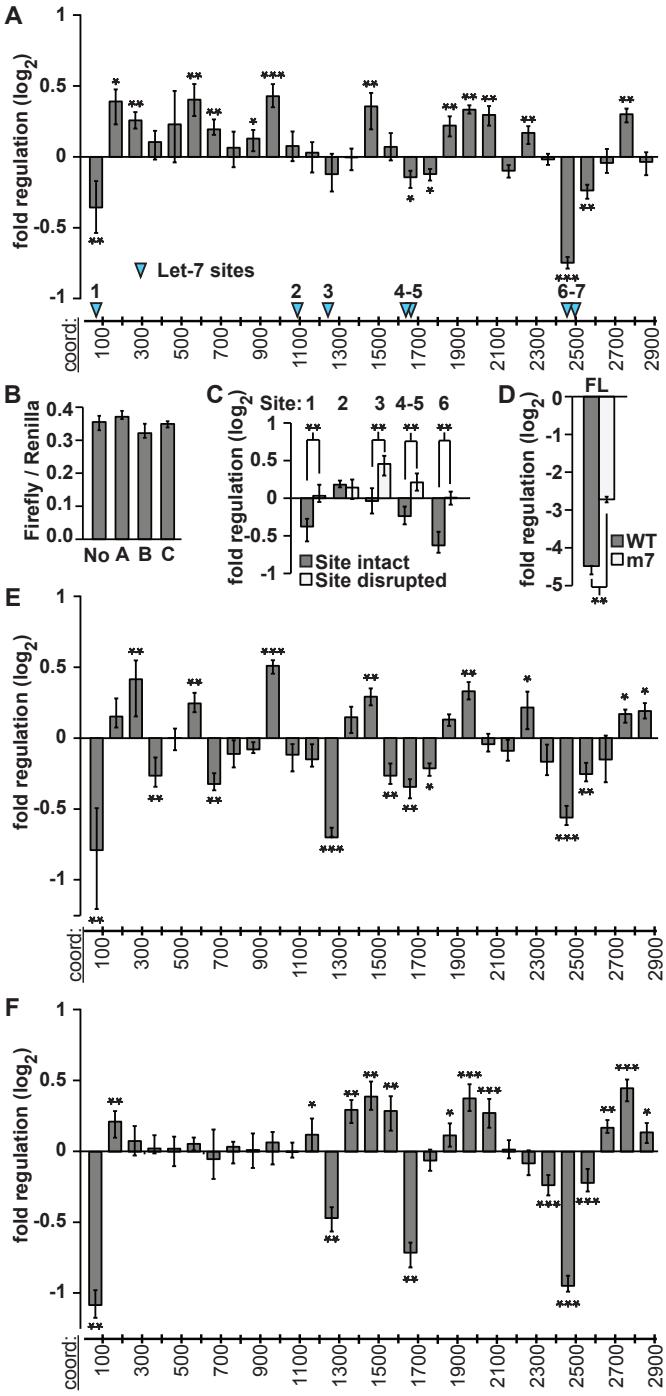
To begin to understand the regulatory sequences present within the *Hmga2* 3′ UTR, we generated a set of luciferase reporter constructs, each containing ~100-nt fragments (100mers) of the mouse 3′ UTR, such that the complete 2.97-kb sequence was represented in the set (Fig. 2.1A). Experiments in which reporter activities are compared with each other can be used to calculate the relative impact on gene expression for each reporter. They cannot, however, be used to infer the absolute impact on gene expression, because there is no way to establish a baseline. To address this problem, we reasoned that 3′ UTRs comprising randomized sequences are unlikely to contain any sequence elements that could impact gene expression, and are therefore ideal inert controls with which to establish a baseline for expression. We generated three different reporter constructs, each

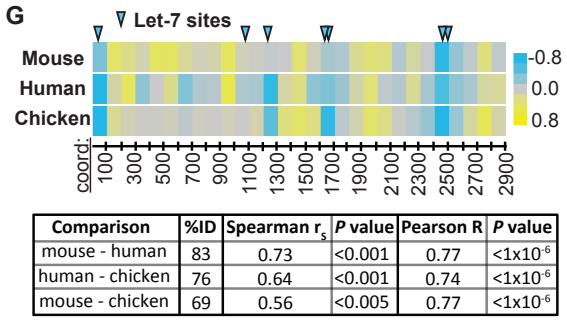
containing a different random 100-nt sequence (random 100mers) as their 3′ UTR. Importantly, there were no significant differences in reporter activity between the three random 100mer constructs, nor between them and a no-3′-UTR control (Fig. 2.1B), indicating that none contained sequences impacting gene expression in our assay and confirming their usefulness as inert controls. Although the random-sequence controls are equivalent in their expression to a no-3′-UTR control for 3′ UTRs of this size, such size-matched random-sequence constructs have the benefit of controlling for the influence of 3′ UTR size *per se* on gene expression [174, 147, 175], a consideration that is more important for larger 3′-UTR fragments (Fig. 2.2A). This approach allowed us to establish the baseline for all of our reporter experiments, thus enabling us to calculate the absolute impact on gene expression for all 3′-UTR fragments assayed.

We first assayed our set of mouse *Hmga2* 100mer reporters in A549 cells (Fig. 2.1A). As expected, the fragments containing previously identified *let-7* miRNA target sites corresponded well to those that significantly down-regulated reporter expression. We confirmed that *let-7* was responsible for the repression observed by comparing reporters with and without intact target sites (Fig. 2.1C). We also measured the total impact of *let-7* repression of the full-length *Hmga2* 3′ UTR by comparing a wild-type full-length construct to a 3′ UTR construct that has all seven *let-7* target sites disrupted (m7) (Fig. 2.1D). *let-7* repression of the full-length matches well with expectation based on the 100mer data. In addition to *let-7* containing fragments, only two other fragments contained active repressive elements, the identities of which were not readily apparent. The presence of such elements, however, is not unexpected, as the *Hmga2* 3′ UTR is considered a strongly repressive sequence (Fig. 2.1D, J; [165]). We observed that many fragments significantly up-regulated reporter expression, a somewhat surprising result given the presence of counteracting repressive elements.

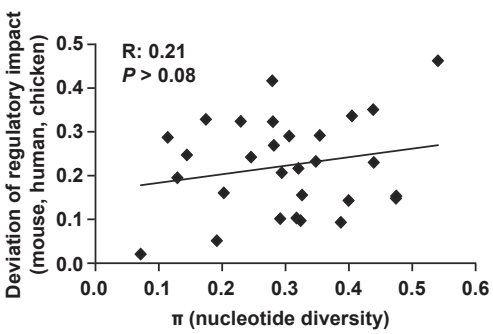
To investigate whether the novel regulatory elements within the *Hmga2* 3′ UTR are

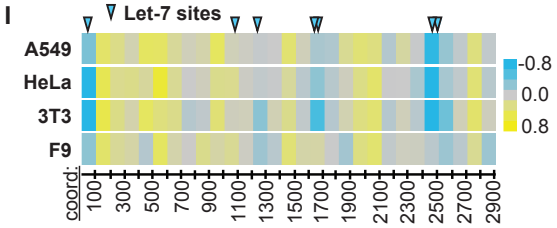
Figure 2.1: Regulatory sequences in vertebrate *Hmga2* 3′ UTRs.
(A) Reporter assays measuring regulatory impact of tiled 100-nt fragments (100mers) of the mouse Hmga2 3 UTR. Histogram indicates $\log_2$-fold change conferred by 3′-UTR 100mers, relative to random-sequence 100mer reporters (B); significance assessed with Bonferroni-corrected Wilcoxon rank-sum tests ($n > 12$; $P < 0.05$, 1×10^{-3}, and 1×10^{-5} significance thresholds indicated by *, **, and ***, respectively; error bars indicate the third largest and third smallest values among 12 replicates, approximating 68% of the data as a nonparametric analog of one standard deviation; for measurements with more than 12 replicates, values were selected so that error bars also approximate 68% of the data). The x-axis shows the approximate coordinates of each 3′-UTR sequence; the positions of *let-7* miRNA target sites are indicated with blue arrowheads. (B) Reporter assays measuring regulatory impact of a no-3′-UTR control (No) and 100mer 3′-UTR fragments of random sequence (A-C). Three different control sequences mediate regulatory effects equivalent to one another and to the no-3′-UTR control ($n = 9$; $P > 0.05$, Wilcoxon rank-sum test). (C) Mutated *let-7* sites abrogate the repression of *Hmga2* 100mer 3′-UTR fragments. Reporter assays comparing the regulatory impact of 100mer fragments containing predicted *let-7* target sites (sites numbered as in A) to otherwise identical fragments in which the *let-7* target sites were disrupted ($n = 9$; ** indicates $P < 1 \times 10^{-3}$, Wilcoxon rank-sum test). (D) *let-7* mediated repression of the full-length *Hmga2* 3′ UTR. As in C, comparing full-length *Hmga2* 3′ UTR constructs (normalized to a no-3′ UTR control) with all *let-7* target sites intact (WT) or disrupted (m7) ($n = 9$, $P < 1 \times 10^{-3}$, Wilcoxon rank-sum test). (E,F) Reporter assays measuring regulatory impact of successive 100-nt fragments of human (E) and chicken (F) *HMGA2* 3′-UTR sequences. Boundaries for human and chicken 3′-UTR fragments are orthologous to mouse coordinates; otherwise as described in A ($n = 9$). (G) Conservation of regulatory sequence impact. Heat maps (top) illustrate reporter data from A,D,E (color key on right); the positions of *let-7* miRNA target sites are indicated with blue arrowheads, as in A. Table (bottom) contains correlation coefficients (Spearman and Pearson) and P values comparing regulation of orthologous 100-nt fragments tiled across the 3′ UTR. (H) Nucleotide diversity (X-axis) is not well-correlated with divergence in regulatory impact of different 3′-UTR fragments from human, mouse, and chicken (Pearson $R = 0.21$, $P > 0.08$). (I) Impact of *Hmga2* 3′-UTR regulatory sequences compared in different cell types. Reporter assays (data depicted as heat maps, as described above) comparing regulatory impact of mouse 100mer 3′-UTR reporters performed in indicated cell lines. Each pair-wise comparison between reporter data is summarized with Spearman and Pearson correlation coefficients along with P values. (J) Regulatory effect of full-length *Hmga2* 3′ UTRs compared in different cell types. Full-length *Hmga2* 3′-UTR constructs (normalized to a no-3′-UTR control) with all *let-7* target sites intact (WT) or disrupted (m7) in different cell types ($n = 9$; * and ** indicate $P < 0.05$ and $P < 1 \times 10^{-3}$, respectively, Wilcoxon rank-sum test).



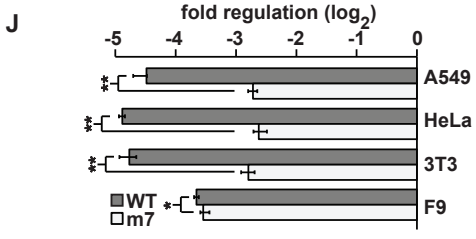


Comparison	%ID	Spearman r_s	*P* value	Pearson R	*P* value
mouse - human	83	0.73	<0.001	0.77	$<1x10^{-6}$
human - chicken	76	0.64	<0.001	0.74	$<1x10^{-6}$
mouse - chicken	69	0.56	<0.005	0.77	$<1x10^{-6}$





Comparison	Spearman r_s	*P* value	Pearson R	*P* value
A549 - HeLa	0.88	$<1x10^{-6}$	0.89	$<1x10^{-6}$
A549 - 3T3	0.91	$<1x10^{-6}$	0.91	$<1x10^{-6}$
A549 - F9	0.62	<0.001	0.61	<0.001
HeLa - 3T3	0.88	$<1x10^{-6}$	0.91	$<1x10^{-6}$
HeLa - F9	0.71	<0.001	0.64	<0.001
3T3 - F9	0.53	<0.005	0.58	<0.001



conserved, we performed an equivalent study of both the human and chicken HMGA2 3′ UTRs (83% and 76% identical to mouse, respectively), in which we generated and characterized a complete series of 100mer fragments in a luciferase reporter system (Fig. 2.1E-G). When generating the human and chicken reporters, the 100mer coordinates were selected based on an alignment of the three 3′ UTRs, allowing us to ensure that comparisons between species were specific to orthologous regions of each 3′ UTR. Importantly, because of the high pair-wise identity, we could confidently identify orthologous positions between the 3′ UTRs. We found significant correlations in regulatory impact of equivalent 3′-UTR fragments between the three species (Fig. 2.1G), indicating conservation of regulatory sequences. The extent to which regulatory impact deviated between species was not strongly correlated to the sequence divergence (nucleotide diversity) of the corresponding 100mer fragments (Fig. 2.1H), perhaps suggesting the existence of relatively small conserved regulatory elements that are responsible for the regulatory effect of each fragment. Notably, both human and chicken HMGA2 3′ UTRs also contained multiple positive regulatory sites, providing additional support for their biological importance within the *Hmga2* 3′ UTR.

To gain additional perspective on the regulatory potential of the *Hmga2* 3′ UTR in a variety of different *trans*-factor environments, we repeated our experiments, testing mouse 100mers in a series of different cell lines: mouse 3T3 and F9 cells and human HeLa cells. Although we found instances of cell-type specific differences in regulation, the overall regulatory landscape was extremely consistent between different cell lines (Fig. 2.1I). The largest exception to this trend was found in F9 cells, a mouse testicular teratoma cell line that should not express *let-7* [176]. Indeed, when we compared the full-length wild-type and *let-7* disrupted (m7) reporters in these cell lines, the F9 cells were the only environment where there was not a substantial difference in expression between the constructs (Fig. 2.1J). Importantly, we selected a range of different types of cells, thus, the majority of the regulatory events detected here are relatively robust to differences in cellular

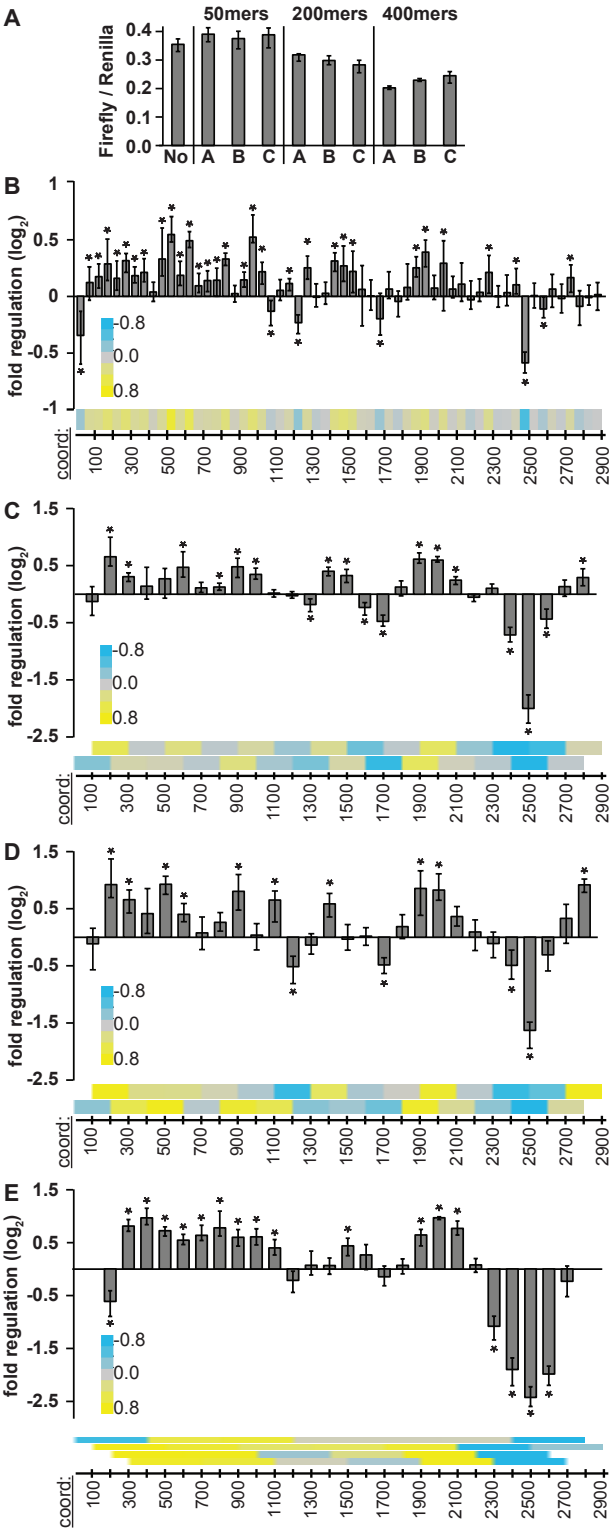
environment.

2.3.2 Map of regulatory sequences within the *Hmga2* 3′ UTR at different nucleotide resolutions

The regulatory capacity of a 3′ UTR cannot be determined by analyzing 100-nt fragments alone, in part because many bona fide regulatory elements are significantly smaller than 100 nucleotides, and also because some elements may require larger surrounding sequence context to recapitulate regulatory impact found in their native context. To address these concerns, we generated three additional sets of reporters, each of different sizes, derived from the mouse 3′ UTR. As before, we generated and assayed size-matched random-sequence control constructs; as expected, random-sequence reporters of the same size impacted reporter expression equivalently (Fig. 2.2A), allowing us to establish an appropriate size-matched baseline for each of the three different sets of reporters. Our observation that longer random-sequence controls were more repressive than shorter ones implies a sequence-independent repressive effect with increasing 3′-UTR length; a result consistent with previous observations [147].

We first generated a 50mer fragment set, which was tiled in nonoverlapping ~50-nt windows across the entire mouse 3′ UTR (Fig. 2.2B), maintaining the same boundaries used previously to generate the 100mer reporters (Fig. 2.1A). Within this high-resolution regulatory map, very few 50mers show repressive effects (6 of 58), and most of these (5 of 6) contain *let-7* target sites, consistent with the 100-nt data set. A large number of fragments (29 of 58) mediate significant positive regulatory effects, again consistent with our analysis of the 100mer reporters. Together, these data indicate that the *Hmga2* 3′ UTR contains at least 35 discrete regulatory elements. It is worth mentioning that the identification of specific regulatory elements within the *Hmga2* 3′ UTR is greatly facilitated by

Figure 2.2: Regulatory sequences in the mouse *Hmga2* 3′ UTR at different size resolutions. (A) Reporter assays measuring regulatory impact of 50-, 200-, and 400-nt ($n = 9$) 3′-UTR fragments of random sequence, otherwise as described in Figure 2.1B. (B-E) Reporter assays measuring regulatory impact of tiled 50mer fragments ($n > 12$) of mouse *Hmga2* 3′-UTR sequences (B); 200mer fragments tiled at 100-nt intervals (C,D, mouse and human sequences, respectively); and 400mer mouse 3′-UTR fragments tiled at 100-nt intervals (E). Heatmaps (bottom of each panel) show same data while illustrating tiling strategy (each bar is centered over the middle of the corresponding square in the heatmap). Significance of regulation determined with Bonferroni-corrected Wilcoxon rank-sum tests ($n = 9$; * indicates $P < 0.05$), otherwise as described in Figure 2.1A.


A
50mers
200mers
400mers
Firefly / Renilla
0.4
0.3
0.2
0.1
0.0
No
A
B
C
A
B
C
A
B
C
B
C
D
E
fold regulation (log2)
1
0.5
0
-0.5
-1
1.5
-1.5
-2.5
-0.8
0.0
0.8
coord:
100
300
500
700
900
1100
1300
1500
1700
1900
2100
2300
2500
2700
2900

narrowing the resolution from 100 to 50 nt (e.g., Fig. 2.6, below).

We next generated a 200mer fragment set, which was tiled in overlapping ~100-nt windows across the entire 3′ UTR using the same boundaries as the original set of 100mers. We generated 200mer sets for both the mouse (Fig. 2.2C) and human (Fig. 2.2D) sequences. Multiple *Hmga2* 200mer fragments also show significant up-regulatory potential and the overall pattern is highly similar between the mouse and human (Pearson correlation coefficient, $R = 0.88$, $P < 110^{-6}$). Finally, we generated an overlapping 400mer fragment set for the mouse 3′ UTR, tiled in ~100-nt offsets using the same boundaries as both the 100 and 200mer sets (Fig. 2.2E). Together, these data sets confirm the presence of multiple positive regulatory sequences within the *Hmga2* 3′ UTR.

It is worth noting that despite the presence of multiple positive sites, the complete *Hmga2* 3′ UTRs of all three species examined are repressive, when compared with a no-3′ UTR control (22-, 12-, and 29-fold repressed for mouse, human, and chicken, respectively). Moreover, the mouse *Hmga2* 3′ UTR remains repressive even when *let-7* target sites have been disrupted (Fig. 2.1J). While some, perhaps the majority, of this repression may derive from the 3′ UTRs large size (top 15% for 3′ UTR length in humans), it may also indicate that the regulatory impact of the full-length 3′ UTR is not simply the sum of its component elements.

2.3.3 Nonadditive interactions between neighboring sequence elements are rare within the *Hmga2* 3′ UTR

Though rare examples of sequence element interactions have been identified, it is not clear how commonly this occurs in mammalian 3′ UTRs. A major motivation for this study was, therefore, to systematically examine a single 3′ UTR for evidence of nonaddi-

tive interactions between different regulatory elements. Since the *Hmga2* 3′ UTR is both long and contains many regulatory sequences (as we have shown above), it is an ideal subject to test how often distinct sequence elements interact. One approach to this question is to compare the regulatory impact of each 3′ UTR fragment to that predicted by the corresponding pairs of smaller fragments. Toward this goal, we modeled the expected regulation of a particular 3′ UTR fragment as the product (log additive) of the regulatory impact of constituent smaller fragments (Fig. 2.3A). This model assumes that sites act autonomously of one another, each contributing independently to the cumulative regulation observed when sites are combined within a 3′ UTR. Notably, this model recapitulates well experimental measurements of the regulatory impact of multiple miRNA target sites [59]. If the observed and the modeled values correlate strongly, this indicates that most sites within the *Hmga2* 3′ UTR function independently of each other. On the other hand, a failure to correlate could indicate the presence of interacting sequence elements, either synergistic or inhibitory.

We implemented the approach described above at three different resolutions, comparing observed to expected reporter data for 100mer, 200mer, and 400mer fragments using, respectively, 50mer, 100mer, and 200mer reporter data to generate expected values. Strikingly, we found strong and significant correlations between observed and expected values at all three resolutions (Fig. 2.3B-D, for mouse 100mers, 200mers, and 400mers, respectively). Indeed, these correlations were almost as strong as simulated correlations for perfectly independent sites that take into account the degree of experimental noise intrinsic to the experiments (Rmax = 0.95, 0.99, and 0.99 for 100mers, 200mers, and 400mers, respectively).

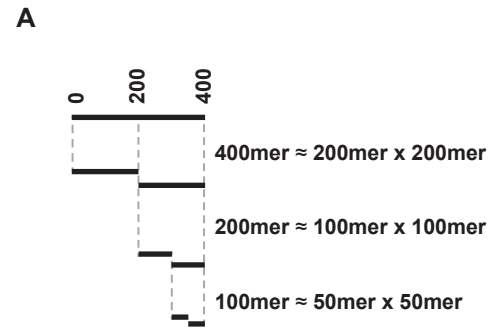
As a more strenuous test of the model of site independence, we tested whether observed 50mer reporter data predicted 200mer and 400mer data, and whether observed 100mer data predicted 400mer data. In all three cases, we observed significant (P

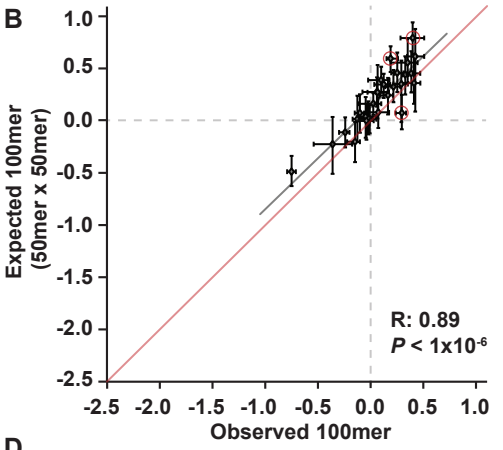
$< 1 \times 10^{-6}$) and pronounced ($R > 0.80$) correlations between the directly measured regulatory impact of large fragments and that derived from measurements of multiple constituent fragments (Fig. 2.3E). These results illustrate that *cis*-element interactions over short to mid-range distances (up to 400 nt) are rare in *Hmga2*, and that the 3′-UTR functions as the sum of many

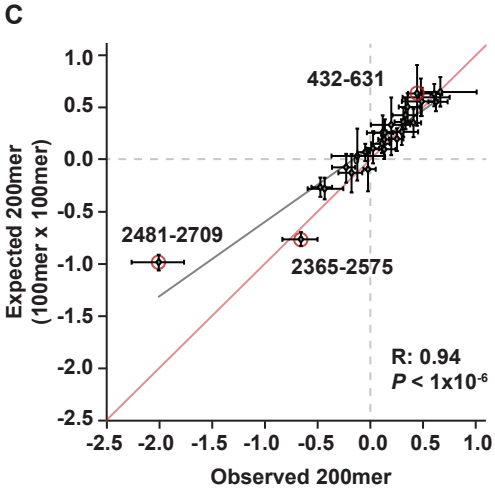
independent elements. Additionally, these results indicate that our approach to mapping regulatory sequences within 3′ UTRs is appropriate, as the vast majority of 3′-UTR fragments we examined consistently recapitulate the behavior of larger portions of the 3′ UTR. This final point is most striking when considering that 50mer reporter data captures well ($R = 0.80$, $P < 1 \times 10^{-6}$) the regulatory impact measured for 400-nt portions of the *Hmga2* 3′ UTR.

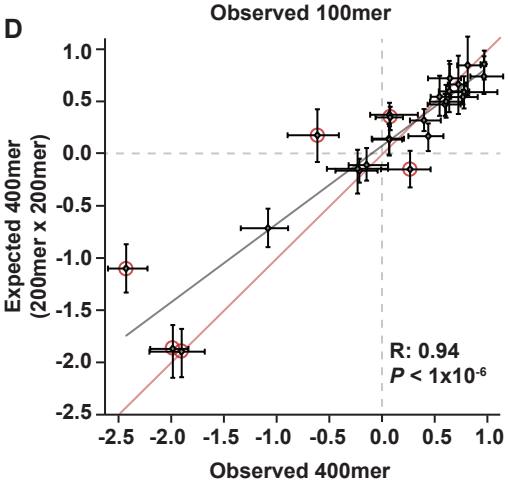
We did observe a small number of fragments with a statistically significant ($P < 0.01$), but usually small, discrepancy between observed and expected measurements (3 of 29 100mers, 3 of 28 200mers, and 6 of 26 400mers). Such exceptions likely derive from two possible sources. Firstly, they may represent bona fide cases in which two (or more) regulatory elements in adjacent fragments interact with either positive or negative synergism. Secondly, they may arise from technical limitations of our approach. For example, a functional regulatory element located in the middle of a 200mer might be divided between each of the constituent 100mers, and therefore no longer functional in either. To differentiate between these possibilities, we mutated sequences at the center of the deviating 200mers (where the two 100mers would meet) and examined the effect on reporter expression (Fig. 2.3F). For two of the three deviating 200mers, this mutational analysis indicated the likely existence of a single regulatory element within the 200mer that is disrupted in both constituent 100mers (these results also likely explain certain of the discrepancies between expected and observed 400mers). In contrast, the remaining 200mer may contain a pair of interacting elements. In summary, only one of 28 200mer fragments (of which

Figure 2.3: 3′-UTR regulatory sequences impact gene expression independently of one another within *Hmga2*.
(A) Depiction of regulatory impact of different 3′-UTR fragments, assuming independent action of regulatory elements. The cumulative effect of regulatory elements within a 3′-UTR fragment is modeled as the product of the regulatory impacts of constituent smaller segments. (B-D) The observed regulatory impact (x-axis) for each 100mer, 200mer, and 400mer fragment compared with the prediction (y-axis) based on the 50mer, 100mer, and 200mer data set, respectively, for the mouse *Hmga2* 3′ UTR. The gray line represents a Pearson best-fit regression between x- and y-values; for comparison, the red line shows a y = x line. Error bars represent the middle 68% of the data as a nonparametric equivalent to one standard deviation. 3′-UTR fragments that deviate significantly (see Materials and Methods) from the regression are circled in red. (E) Comparison of the observed regulatory impact of 200mer and 400mer 3′-UTR fragments to the prediction using values observed for 50mers (for 200mers and 400mers) and 100mers (for 400mers). Each comparison is summarized with correlation coefficients (Spearman and Pearson) and *P* values. (F) Sequences located at the junction of adjacent 100mer fragments alter the regulatory behavior of two of the three 200mer fragments whose activity is not well predicted by constituent 100mers (circled red in C). For each 200mer fragment, mutations were introduced at the junction between the two corresponding 100mers. The activity of each mutant derivative was normalized to the original 200mer reporter constructs. Significant difference between intact and disrupted junction constructs was determined with a Wilcoxon rank-sum test ($n = 9$; * indicates $P < 0.05$).



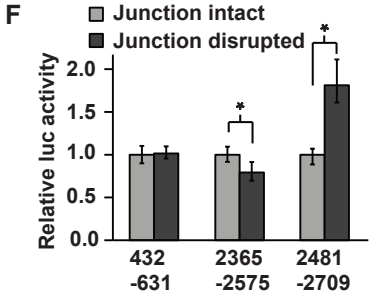






E

Comparison	Spearman r_s	*P* value	Pearson R	*P* value
50-200mer	0.82	$<1x10^{-6}$	0.80	$<1x10^{-6}$
50-400mer	0.85	$<1x10^{-6}$	0.84	$<1x10^{-6}$
100-400mer	0.91	$<1x10^{-6}$	0.95	$<1x10^{-6}$



almost all contain detectable regulatory sequences) showed evidence indicating nonadditive interactions between regulatory elements. Taken together, these data indicate that synergistic interactions between regulatory elements rarely occur within this 3′ UTR, at least within the size resolutions we examined.

2.3.4 The *PIM1* 3′ UTR contains multiple regulatory elements, which largely function independently

Fundamentally, our comprehensive study of the *Hmga2* 3′ UTR identified many discrete regulatory sequence elements with minimal nonadditive interactions between neighboring elements. We tested whether these results extend to an additional 3′ UTR by creating a similar data set for the *PIM1* 3′ UTR, which is relatively well conserved [177] but has not been systematically characterized with respect to post-transcriptional biology. As before, we generated both 100- and 200-nt sets of reporters, tiled at 100-nt intervals across the complete *PIM1* 3′ UTR. Reminiscent of our results with *Hmga2*, most (8 of 13 100mers and 10 of 12 200mers) of the 3′-UTR fragments mediated significant regulation (Fig. 2.4A,B). Notably, the regulatory impact of 200mers was again well recapitulated by constituent 100mers (Fig. 2.4C). Also, as with *Hmga2*, the full-length *PIM1* 3′ UTR is repressive (compared with a no-3′-UTR control; Fig. 2.4D) despite the presence of multiple positive sites, although the extent to which the repression can be attributed to the size of the *PIM1* 3′ UTR could not be determined. Along with our investigations of regulatory sequences within the *Hmga2* 3′ UTR, these results imply that the regulatory sequence elements found in 3′ UTRs typically function independently of one another.

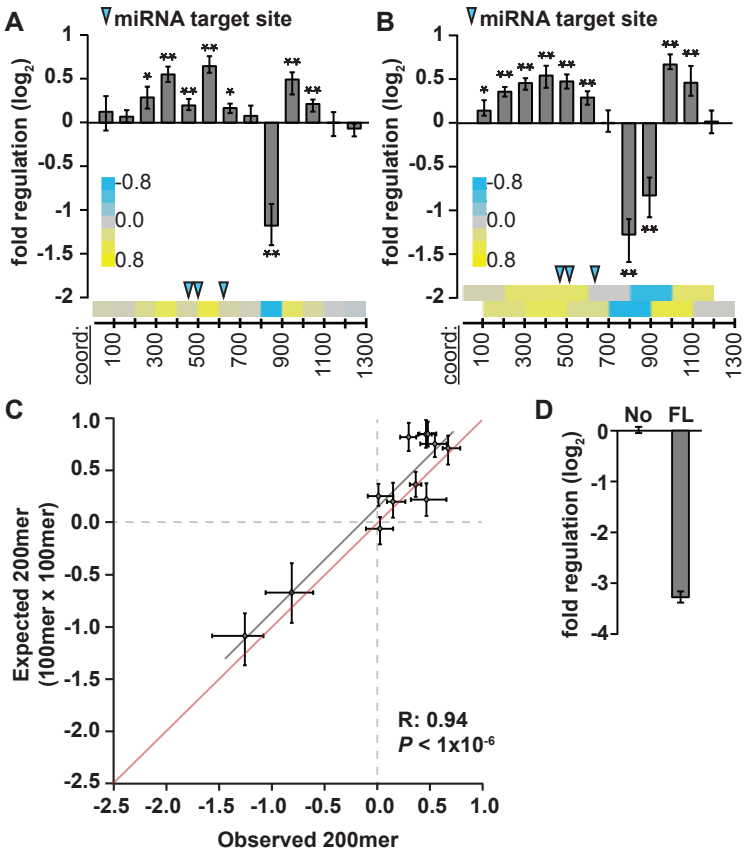


Figure 2.4: Regulatory sequences impact gene expression independently of one another within the *PIM1* 3′ UTR.
(A,B) Reporter assays measuring regulatory impact of 100- and 200-nt 3′-UTR fragments (A,B, respectively) tiled across the PIM1 3′ UTR at 100-nt intervals ($n = 9$; $P < 0.05$ and 1×10^{-3} indicated by * and **, respectively, Bonferroni-corrected Wilcoxon rank-sum tests). Heatmaps (bottom of each panel) show same data while illustrating tiling strategy; blue arrows indicate location of target sites for miRNAs expressed in A549 cells. (C) The measured regulatory impact of each 200mer fragment (observed, x-axis) modeled as the product of the regulatory impact of constituent 100mer fragments (expected, y-axis), for mouse *PIM1* 3′-UTR sequences, otherwise as described in Figure 2.3C. (D) The full-length *PIM1* 3′ UTR (FL) is repressive (normalized to a no-3′ UTR control [No], $n = 9$).

2.3.5 A role for HuR in mediating regulation within the *Hmga2* 3′ UTR

When examining the up-regulating *Hmga2* 50mer sequences, we noted that many contained U-rich sequences consistent with a possible function as AU-rich elements (AREs). AREs interact with multiple different *trans*-factors and their presence within a transcript can mediate both positive and negative post-transcriptional regulatory effects (for review,

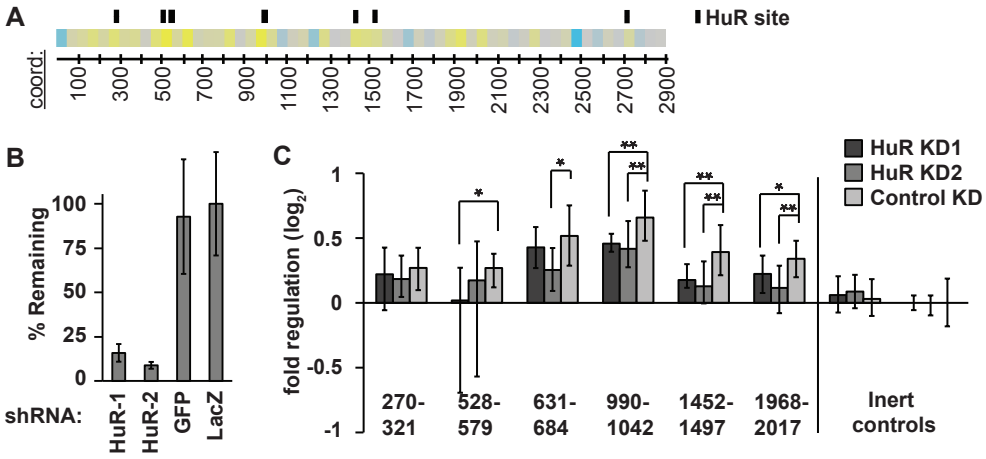


Figure 2.5: A role for HuR in regulation of *Hmga2*.
(A) HuR binding sites correspond to up-regulating 50mers. Location of HuR binding sites (black boxes) according to available PAR-CLIP data [180] displayed parallel to a heatmap representation of the *Hmga2* 3′-UTR 50mer reporter data. (B) Evaluation of HuR knockdown. A549 cells were infected with two different shRNA hairpins targeting the HuR mRNA, and the effect on HuR transcript levels determined with qRT-PCR; two different hairpins (targeting GFP and LacZ) were used as negative controls ($n = 3$). Error bars indicate one standard deviation. (C) Reporter assays of selected *Hmga2* 50mer reporters, and one random 50mer control reporter, in HuR knockdown cells. Reporter expression, normalized to a second inert random-sequence control (rightmost data), was compared between HuR knockdown cells and cells infected with control shRNAs (targeting GFP). Significant changes in reporter expression were determined using Wilcoxon rank-sum tests ($n = 12$; $P < 0.05$ and 1×10^{-3} indicated by * and **, respectively).

see [63]. HuR is an established ARE-binding protein and is one of the few known to confer positive regulation [178, 179]. Publicly available PAR-CLIP data [180] indicate several HuR binding sites within the *Hmga2* 3′ UTR. Moreover, all such sites fall within fragments we identified as containing positive regulatory elements (Fig. 2.5A). To test whether HuR is required for the observed positive regulation of the *Hmga2* 3′-UTR fragments, we used RNAi (using lentiviral-delivered shRNAs) to knock down HuR in A549 cells. The efficacy of HuR knockdown was assessed by qRT-PCR (Fig. 2.5B), identifying two different shRNAs that effectively target HuR. We then performed reporter assays in the HuR knockdown cells to determine the effect on *Hmga2* 3′-UTR reporter constructs. Inert shRNAs targeting GFP and LacZ served as negative controls for RNAi.

We found consistent evidence that HuR is required for the positive regulatory impact of multiple 50mer reporter fragments (Fig. 2.5C). For three different 50mer fragments, both HuR knockdown experiments resulted in significant reductions in positive regulatory impact. For the remaining three 50mer fragments, HuR knockdown with either shRNA consistently impaired the positive regulatory impact of the reporters, but only two of the experiments were statistically significant. Overall, these results suggest a role for HuR in regulating the *Hmga2* transcript. Surprisingly, however, the results did not extend to a full-length *Hmga2* 3′-UTR reporter, as it was only minimally affected by HuR knockdown (data not shown; see Discussion).

2.3.6 Mapping positive regulatory sequence elements within the *Hmga2* 3′ UTR

To further examine the positive regulatory impact conferred by fragments of the *Hmga2* 3′ UTR, we performed a scanning mutagenesis analysis of the three 50mer fragments with the strongest positive impact on reporter expression. We first generated a series of reporter constructs in which we replaced 12 nt of endogenous 3′-UTR sequence with 12 nt of inert sequence (derived from our random-sequence controls), with the 12-nt window tiled at 8-nt intervals across each of the three 50mers. We then assayed the reporters, expecting to identify a subset whose ability to mediate increased reporter activity was compromised, thereby identifying the specific nucleotides comprising the positive regulatory elements within these *Hmga2* fragments. This strategy was successful for two of the three 50mer reporters (Fig. 2.6B,C), identifying nucleotides 660-671 and 1009-1029 as the specific sequences required for positive regulation. For the remaining 50mer the results were less clear, possibly indicating the existence of multiple separate regulatory elements (Fig. 2.6A).

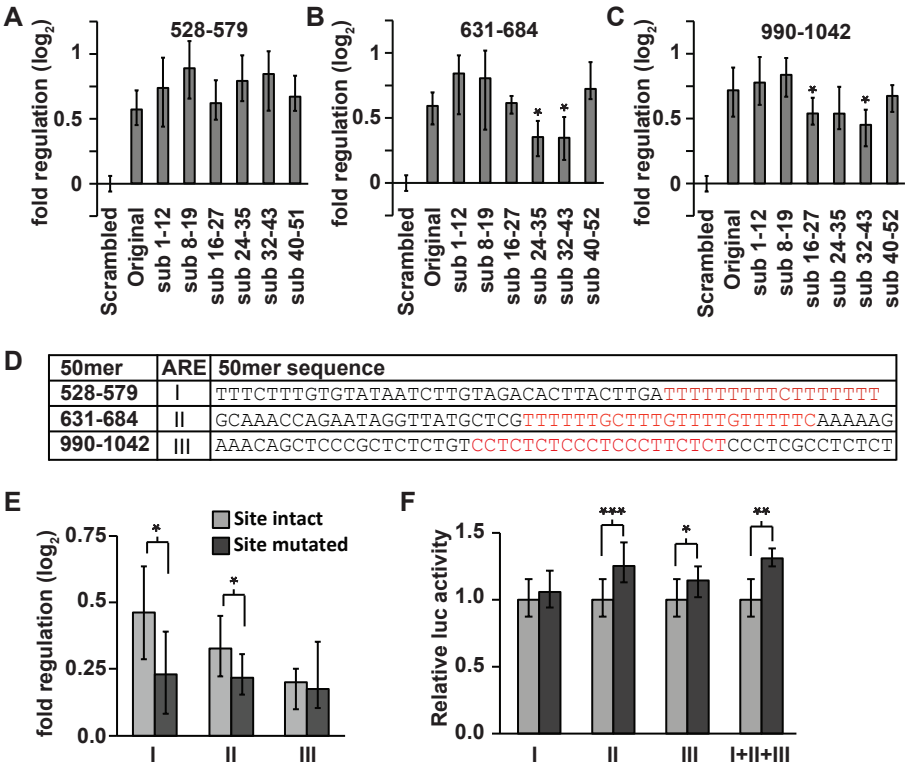


Figure 2.6: Fine-resolution mapping of positive regulatory sequence elements within the *Hmga2* 3′ UTR.

(A-C) Scanning mutagenesis identifies regulatory sites within 631-684 and 990-1042 50mers (B,C, respectively), but not within 528-579 50mer (A). Reporter assays comparing the original endogenous-sequence 50mers to substitution mutant 50mers, relative to size-matched inert random-sequence controls. Significant reductions (Wilcoxon rank-sum tests) in positive regulatory impact are signified with an asterisk ($n = 9$). (D) Sequence of three strongest positive 50mers with candidate sequence elements highlighted in red. ARE sequence elements are denoted I, II, and III. (E,F) Reporter assays determining the effect of a targeted deletion of candidate elements from 200mers (E) or full-length Hmga2 (F). Reporter activities of 200mers are shown relative to size-matched random-sequence inert controls; full-length reporters with the targeted deletion are shown relative to an intact *Hmga2* 3′-UTR reporter. Significant effects on reporter expression were determined using Wilcoxon rank-sum tests ($n \geq 12$; $P < 0.05$, 1×10^{-3}, and 1×10^{-5} significance thresholds indicated by *, **, and ***, respectively).

The sequences identified by the scanning mutagenesis as candidate positive regulatory elements coincided well with U-rich or CU-rich sequences (consistent with AREs) (Fig. 2.6D). Indeed, even the 50mer refractory to this approach (Fig. 2.6A) contains a similar pyrimidine-rich sequence (Fig. 2.6D). To determine if U-rich sequences are responsible for these up-regulating effects, we deleted the candidate AREs from 3′-UTR reporters containing either the full-length *Hmga2* 3′ UTR or 200mer fragments containing the U-rich sequences (Fig. 2.6D-F). Experiments performed with 200mer fragments were consistent with the mutational analyses, supporting a role for pyrimidine-rich tracts as positive regulatory sequences (Fig. 2.6E). Surprisingly, removal of poly-pyrimidine tracts within the full-length *Hmga2* had an effect of the same magnitude but in the opposite direction (Fig. 2.6F), suggesting that the same sequences can also function as negative regulatory elements, but only when located within the complete 3′ UTR. Thus, while it was clear that these sites regulate expression, the exact nature of this regulation is remarkably context dependent.

2.3.7 Terminal sequences within the *Hmga2* 3′ UTR induce a functional switch in candidate AREs from positive to negative regulatory elements

As individual AREs are known to bind many different proteins (for review, see [63]), it is conceivable that events elsewhere in the 3′ UTR can modulate which ARE-binding protein stably associates with AREs within that same 3′ UTR, resulting in a switch in function. We therefore hypothesized that sites elsewhere in the *Hmga2* 3′ UTR were responsible for the functional switching of the candidate AREs responsible for up-regulation within 200mer fragments to down-regulation in the full-length *Hmga2* 3′ UTR (Fig. 2.6). To identify the region responsible, we generated a truncation series of the full-length *Hmga2*

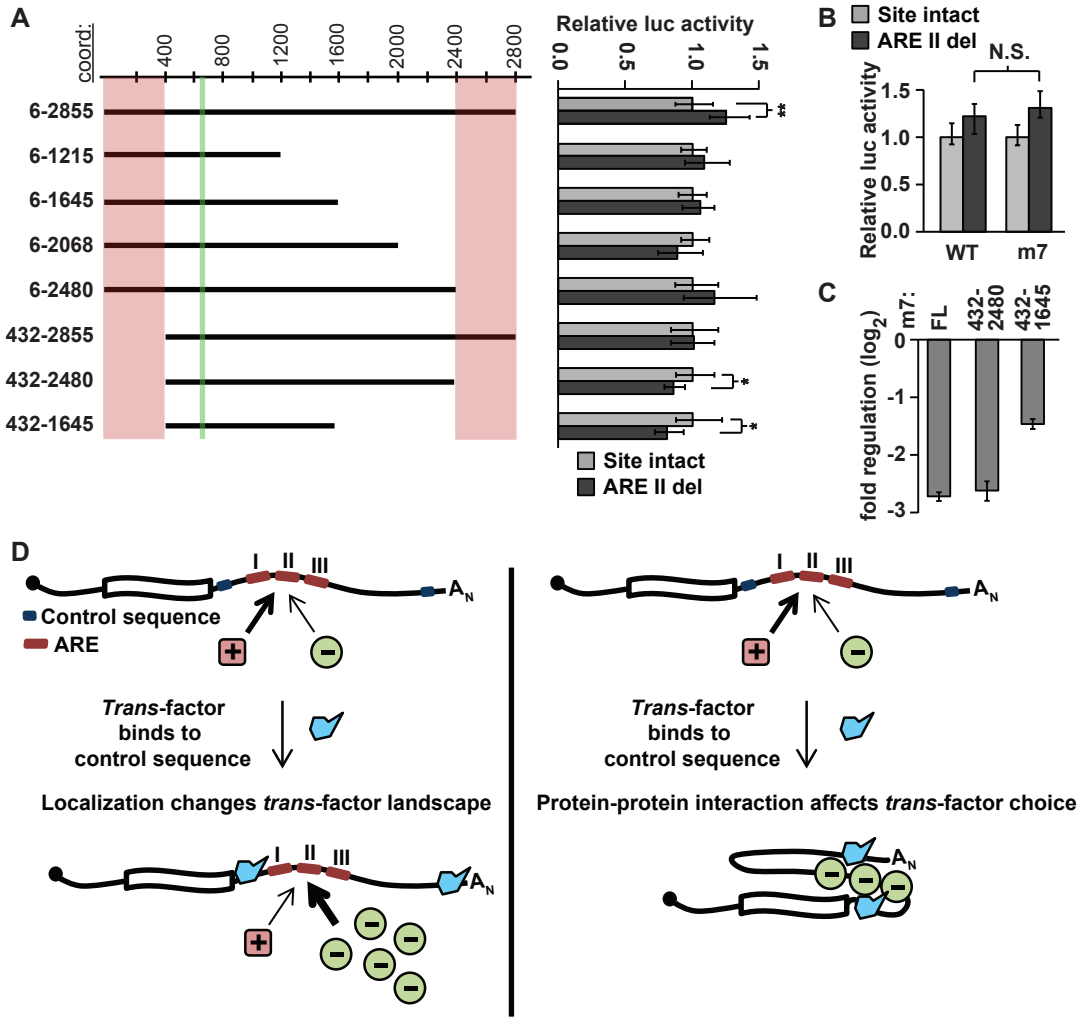
3′ UTR, and compared this series with a parallel truncation series in which the ARE at nucleotides 660-671 (ARE II) was deleted. These experiments revealed two regions of the 3′ UTR that alter the behavior of the distal ARE: nucleotides 6-431 and 2481-2855 (Fig. 2.7A). In 3′-UTR reporters lacking either region, the ARE is rendered inert. Furthermore, we found that for 3′-UTR reporters lacking both regions, the same ARE functions as a positive regulatory element. Any additional truncation had no effect on the function of the ARE sequence. This pattern was recapitulated when tested in the mouse 3T3 cell line (data not shown). Together, these results indicate that distal sequences within the 3′-UTR control the function of at least one ARE located hundreds of nucleotides distant. As individual AREs are known to bind many different proteins (for review, see [63]), it is conceivable that events elsewhere in the 3′ UTR can modulate which ARE-binding protein stably associates with AREs within that same 3′ UTR, resulting in a switch in function. We therefore hypothesized that sites elsewhere in the *Hmga2* 3′ UTR were responsible for the functional switching of the candidate AREs responsible for up-regulation within 200mer fragments to down-regulation in the full-length *Hmga2* 3′ UTR (Fig. 2.6). To identify the region responsible, we generated a truncation series of the full-length *Hmga2* 3′ UTR, and compared this series with a parallel truncation series in which the ARE at nucleotides 660-671 (ARE II) was deleted. These experiments revealed two regions of the 3′ UTR that alter the behavior of the distal ARE: nucleotides 6-431 and 2481-2855 (Fig. 2.7A). In 3′-UTR reporters lacking either region, the ARE is rendered inert. Furthermore, we found that for 3′-UTR reporters lacking both regions, the same ARE functions as a positive regulatory element. Any additional truncation had no effect on the function of the ARE sequence. This pattern was recapitulated when tested in the mouse 3T3 cell line (data not shown). Together, these results indicate that distal sequences within the 3′-UTR control the function of at least one ARE located hundreds of nucleotides distant.

Notably, the 3′-UTR regions that affect the distant ARE both contain highly effective *let-7* target sites (Fig. 2.1A; [168]), suggesting that *let-7* targeting might play a role in this

phenomenon. To address this possibility, we deleted ARE II from a full-length *Hmga2* 3′-UTR construct that had all seven *let-7* target sites disrupted. Orthogonally, we also examined our truncation reporters (Fig. 2.7A) in F9 cells, which do not express *let-7*. If *let-7* targeting is necessary for the ARE to convert to a repressive element, this conversion should occur in neither F9 cells nor in reporters containing disrupted *let-7* target sites. Instead, our data clearly indicates that loss of *let-7* targeting (through loss of either the target site or the miRNA; Fig. 2.7B and data not shown, respectively) had no effect on the functional switch in the ARE. Thus, the *Hmga2* 3′ UTR contains unknown terminal regulatory elements whose function appears to render ineffective internal positive regulatory elements, converting them to repressive elements.

To investigate whether terminal regulatory elements induce a more global switching of otherwise positive regulatory elements within the *Hmga2* 3′ UTR, we directly compared the full-length sequence to reporters containing terminal deletions. In particular, we were interested in whether such experiments could explain the apparent discrepancy between the repressiveness of the full-length 3′ UTR and the prevalence of positive regulatory elements detected in 50-400-nt reporters. To better enable comparisons between the different reporters, each of which contain different numbers of *let-7* target sites, we tested only reporters in which the *let-7* target sites were disrupted (m7). Although a reporter containing nucleotides 432-2480 of the *Hmga2* 3′-UTR-mediated regulation equivalent to that of the full-length reporter, the activity of a reporter containing nucleotides 432-1645 was significantly and substantially increased ($P < 1 \times 10^{-4}$; ~2.4-fold), compared with the full length (Fig. 2.7C). These results support a model whereby the deleted sequences normally act as control sequences involved in distal element switching. However, it is important to note that such experiments are unable to distinguish between regulation conferred by specific regulatory elements, and the effect of 3′-UTR size, *per se* (Fig. 2.2A; Hogg and Goff 2010; Nicholson et al. 2010). In particular, the results highlight sequences within nucleotides 6-431 and 1645-2844 as important for the complete repressive effect

Figure 2.7: Mapping regions involved in switching ARE function.
(A) Truncation analysis reveals regions necessary for switching. The effect on reporter expression of ARE II deletion (shown on right) was determined in a set of truncation reporter constructs (illustrated on left with approximate coordinates shown at top). For each pair of truncation constructs, significant differences between otherwise wild-type and ARE II-delete reporters were determined with Wilcoxon rank-sum tests ($n = 9$; $P < 0.05$ and 1×10^{-3} indicated by * and **, respectively). The location of ARE II is marked in green and regions that affect ARE II function are marked in pink. (B) ARE II deletion has an equivalent effect on reporter expression in both a wild-type and a *let-7*-disrupted full-length *Hmga2* reporter (Wilcoxon rank-sum tests; $n = 9$, $P > 0.2$). (C) Impact of terminal deletions on the overall regulatory effect of the *Hmga2* 3′ UTR. Full-length *Hmga2* 3′ UTR (m7) is compared with terminal truncation mutants (all constructs normalized to a no-3′ UTR control) ($n = 9$). (D) A model showing two possible mechanisms for the ARE II switching phenotype. On the left, a localization model depicts a trans-factor (in blue) binding to localization signals within the terminal regions, leading to a change in the transcripts local environment, thereby altering the set of *trans*-factors available to bind ARE II. On the right, a protein-protein interaction model depicts a *trans*-factor (in blue) interacting with sequences within the terminal regions of the 3′ UTR; binding of this factor governs the selection of regulatory proteins bound to ARE II.


A
coord:
400
800
1200
1600
2000
2400
2800
6-2855
6-1215
6-1645
6-2068
6-2480
432-2855
432-2480
432-1645
Relative luc activity
0.0
0.5
1.0
1.5
#
*
*
Site intact
ARE II del
B
Site intact
ARE II del
N.S.
Relative luc activity
1.5
1.0
0.5
0.0
WT
m7
C
m7:
FL
432-2480
432-1645
fold regulation (log2)
0
-1
-2
-3
D
I
II
III
AN
Control sequence
ARE
Trans-factor binds to control sequence
Localization changes trans-factor landscape
Trans-factor binds to control sequence
Protein-protein interaction affects trans-factor choice

observed for the full-length 3′ UTR.

2.4 Discussion

Prior to this study, the most widely studied aspect of HMGA2 regulation was the post-transcriptional repression of the *Hmga2* transcript by *let-7* [167, 168]. Notably, however, *let-7* was responsible for only ~15% of the total regulatory effects observed in our study. In fact, our findings illustrate that many previously uncharacterized regions, distributed throughout the *Hmga2* 3′ UTR, have significant regulatory impactmost of comparable magnitude to the *let-7* target sites. The majority of 3′-UTR fragments mediated similar effects when compared between mouse, human, and chicken sequences, a level of conservation likely indicating important biological roles. Moreover, the regulatory map of the 3′ UTR was remarkably robust to different cellular environments, suggesting that the corresponding *trans*-factors are broadly expressed.

In the study that originally determined that *Hmga2* expression was regulated by sequences within its 3′ UTR, Borrmann et al. characterized it using a strategy based on successive ~500-nt truncations [165]. Their analysis revealed two negative and a single positive regulatory region, which, while informative, gives an incomplete depiction of the complexity of post-transcriptional regulation of *Hmga2*. Given the relatively small size of many 3′-UTR regulatory elements and the absence of reliable predictive tools, high-resolution mapping is one of the few appropriate approaches to defining a near-complete set of regulatory sequences within a 3′ UTR. Here, by using this approach, we identify at least 35 distinct regulatory elements within the *Hmga2* 3′ UTR, a density of ~12 elements per 1000 nt of 3′-UTR sequence. Given the biological importance of post-transcriptional control of HMGA2 by *let-7* [167, 168], these newly identified *cis*-regulatory regions may also play important roles in development and oncogenesis, through their

control of HMGA2 levels.

The extent to which discrete *cis*-regulatory elements in 3′ UTRs act synergistically (either positively or negatively) is a major unanswered question in post-transcriptional biology. Here, we attempted to answer this question using a methodical search for evidence of synergism within the *Hmga2* and *PIM1* 3′ UTRs. Although our comprehensive analysis was limited to two 3′ UTRs, in both cases our results indicate that such interactions, at least over the distance ranges we examined (up to 400 nt), are not a typical paradigm for *cis*-elements within 3′ UTRs. Our finding that most sites act independently, if generalizable, will greatly simplify endeavors toward a full understanding of the impact of 3′ UTRs on gene expression. However, our data also indicate that despite an absence of synergistic interactions over short to medium distances, the behavior of sites in the full-length 3′ UTR is not necessarily fully recapitulated by behavior of sites within smaller fragments.

In many cases, synergism can derive from local effects, whereby binding of a *trans*-factor at one site affects the affinity of a second *trans*-factor to a neighboring site. This can be accomplished through changes in RNA conformation [111] or through physical interactions between *trans*-factors, resulting either in increased affinity to sites [181] or blocking of a second site through steric hindrance [118]. These types of interactions are discoverable by our approach, although we may not capture every instance. Another, perhaps more rare, mode of synergism derives from long-distance interactions. An example would be a site at one end of a 3′ UTR that controls transcript localization (for example to processing bodies); such a change in localization can readily alter the set of *trans*-factors available to the transcript and thereby alter the function of a *cis*-element anywhere within a 3′ UTR. Our primary approach is not well suited to identify such interactions, as the maximum fragment size we examined was 400 nt. It is clear that long-range interactions do occur; indeed IGF2BP3 control of the HMGA2 transcript is one such example

[169], as is our discovery of distal control of AREs within the *Hmga2* 3′ UTR. Systematic identification of such interactions requires both knowledge of the regulatory impact of discrete pieces of a 3′ UTR, together with experiments that investigate the roles of identified elements within the complete native 3′ UTR. It remains to be seen how frequent such complex interactions are. It is worth mentioning that even if such interactions are rare, their impact on regulation can be profound, and may even be relatively common in a small subset of 3′ UTRs whose post-transcriptional regulation is particularly critical.

Although the primary focus of this study was to characterize *cis*-regulatory elements in the *Hmga2* 3′ UTR, we also showed that disrupting the expression (by RNAi) of the positive regulatory RBP HuR had a significant effect on multiple individual fragments of the *Hmga2* 3′ UTR. Surprisingly, we were unable to see the same effect on the full-length 3′ UTR. While this may indicate that HuR has a minimal role in *Hmga2* regulation, there are other possible explanations. The expression of RBPs that bind AREs (ARE-BPs), including HuR, are known to be extensively interconnected, such that inhibition of one ARE-BP results in complex changes in the levels of other ARE-BPs [182]. For example, repression of HuR led to at least twofold changes in the protein levels of three other ARE-BPs, with KSRP and TIA-1 levels going down and AUF1 going up. Thus, knockdown of HuR may yield interpretable results when focusing on isolated, discrete 3′-UTR fragments that contain only AREs with high affinity to HuR, but results may not be so clear for a 3′-UTR sequence containing multiple AREs, some of which can bind other ARE-BPs. A recent study identified another RBP that has an important role in HMGA2 regulation: Jnson et al. demonstrated that IGF2BP3 protects *let-7* targeted transcripts, and HMGA2 in particular, from miRNA repression by sequestration into IGF2BP3 granules" [169]. Unfortunately, we were unable to address this aspect of *Hmga2* biology, as in our hands, shRNA-mediated inhibition of IGF2BP3 had no effect on full-length *Hmga2* 3′-UTR reporters (data not shown). This has two likely explanations: It is possible that a stronger knockdown is needed to see an effect, or, it may be that IGF2BP3 does not regulate *Hmga2*

in A549 cells.

Although our results indicate that interactions between 3′-UTR elements are rare, we did discover one striking exception within *Hmga2*. In isolation, our data clearly indicate that AREs within the *Hmga2* 3′ UTR are positive regulatory elements, but in the context of the full-length sequence, the same AREs became repressive elements. We demonstrated that terminal sequences within the 3′ UTR are responsible for the conversion of at least one of the internal AREs from activating to repressive sequence elements. We also showed that this effect was fully independent of *let-7* targeting. The most likely explanation for this switch in behavior is a change in the identity of the ARE-BP associated with the site, from HuR to an ARE-BP that mediates repression. We have considered three possible mechanisms that could be driving this interaction. The first model, and the one we favor, is that sub-cellular localization of the *Hmga2* transcript is controlled by sequences within the terminal regions of the 3′ UTR; altered localization, in turn, changes the RBPs available for binding to *Hmga2* (Fig. 2.7D, left). The second model posits that physical interactions between *trans*-factors bound at terminal sequences within the 3′ UTR and ARE-binding proteins control the identity of the specific ARE-BP that binds to the central sites (Fig. 2.7D, right). The final model is that the native function of the ARE elements are determined by the structural conformation of the *Hmga2* 3′ UTR, which requires the sequences at both termini. This final possibility seems least likely for two reasons: We can find no plausible predicted RNA structures supporting this model, and terminal truncations at only one end inactivated the internal ARE but did not induce switching. Determining which of these mechanisms is responsible will require further study. While we only demonstrated the switching effect for the three strongest positive regulatory sites (and only mapped the interacting region for one of them), it is possible, indeed likely, that this applies to many of the remaining sites we identified as positive regulatory elements. If true, this may explain the seemingly contradictory identification of multiple positive sites that would counteract the negative effect of *let-7* targeting, as well as the ap-

parent contrast between the regulatory effect of the full-length 3′ UTR and the pieces that comprise it. It is easy to imagine how such switchable AREs could contribute to regulation of *Hmga2*. In adult tissues, when repression of *Hmga2* is essential for normal cellular function, the AREs function as repressive sites, along with *let-7* target sites. However, in situations where *Hmga2* is highly expressed, such as certain embryonic tissues, the same AREs could act as positive regulators of *Hmga2* expression.

In summary, our study has three major findings. First, the *Hmga2* 3′ UTR is crowded with regulatory sequences, many of which were previously unidentified. Second, despite the density of regulatory sequences, synergistic interactions over short to medium distances are exceedingly rare; we identified and validated only one such example, and found that all other adjacent regulatory sequences largely functioned independently. These findings extended to a second 3′ UTR, indicating the discoveries may be generally applicable. If true, the lack of widespread interactions means that useful information can be gathered by studying sequence elements in isolation with limited surrounding sequence context. However, it is clear that complex interactions between elements do occur and can have pronounced effects on the role of a 3′ UTR. This is particularly clear in light of our third discovery: A long-range interaction within *Hmga2* reconfigured multiple positive elements to repressive sequence elements. That synergistic interactions appear to be rare within 3′ UTRs suggests that those evolved to contain them likely correspond to genes whose post-transcriptional control is of particular biological importance. The interactions identified to date have commonly involved genes with important roles in cell-cycle control [111], development [118], and cancer [112], processes that are generally under very complex control. Though preferences in research topics may account for some of this apparent enrichment, it would be interesting to see if an unbiased search for element interactions would find interesting patterns in which types of genes evolves interacting 3′-UTR elements.

2.5 Materials and Methods

3′-UTR reporter constructs

Forward and reverse PCR primers were designed at ~100-nt intervals spanning the entire *Hmga2* (mouse, human, and chicken) and *PIM1* (human) 3′ UTRs. Each primer also contained SpeI/NheI (forward) or NotI (reverse) restriction sites to use for cloning. PCR amplicons of ~100 and ~200 (*Hmga2* and *PIM1*) and 400 bp (*Hmga2* only) were generated using these primers, creating amplicons tiled at ~100-nt intervals across each 3′ UTR. To generate the 50-nt insert sequences, oligos were designed so that they could anneal to one another, and after primer extension create double-stranded molecules tailed with appropriate restriction sites. Each 3′ UTR segment was then inserted downstream from the firefly luciferase coding sequence and upstream of the SV40 late poly(A) signal of Promega's pmirGLO Dual-Luciferase miRNA Target Expression vector, using the NheI and NotI vector restriction sites. The insert sequences and cloning junctions of all resulting constructs were fully sequence-validated.

Random-sequence and no-3′-UTR control reporter constructs

All luciferase reporter experiments were performed using a panel of size-matched reporter constructs in which the sequence of the 3′ UTR was randomly generated. The sequence of the random 100mer 3′ UTRs were generated to mimic the base composition of the *Hmga2* 3′ UTR (A:29%, U:30%, C:23%, G:18%). These percentages represent the nucleotide ratios in the full-length *Hmga2* 3′ UTR, and are close to the average composition of all 3′ UTRs (A:27%, U:29%, C:22%, G:22%). Three different random 100mer sequences were generated, and oligonucleotides designed to generate the corresponding double-stranded DNA suitable for cloning into pmirGLO.

Three different 50mer control sequences were derived from 100mer control sequences, and were cloned using oligonucleotide extension, as described above. Three different 200mer control sequences were made by combining different random 100mer controls using overlap extension PCR. Two different random 400mer sequences were designed as described above for random 100mers and one was designed as a combination of two neutral 200mers. All three random 400mers were synthesized as gBlocks (IDT), which were inserted into pmirGLO. The no-3′ UTR control was generated by digesting the pmirGLO vector with NheI and NotI and using T4 DNA polymerase to create blunt ends, which were then ligated.

Mutated 3′-UTR reporters

let-7 mutants (Fig. 2.1C,D,J): 100mer reporters containing disrupted *let-7* target sites were PCR amplified from an *Hmga2* 3′-UTR construct in which all *let-7* binding sites were disrupted [168]; amplicons were cloned into the pmirGLO vector, as described above.

Junction mutants (Fig. 2.3F): For fragments with coordinates 432-631 and 2365-2575, 10 nt of 3′ UTR sequence were removed from the region where the two corresponding 100mers would meet and replaced with scrambled versions of the sequences. For 2481-2709, the *let-7* target site that is located at the junction between the two corresponding 100mers was disrupted by altering three nucleotides within the seed region. Mutants were generated using overlap extension PCR. The altered versions were inserted into pmirGLO, as described above.

Scanning mutagenesis (Fig. 2.6A-C): For each of the three positive 50mers we generated a set of mutants by replacing 12 nt of endogenous 3′-UTR sequence with 12 nt of inert 3′-UTR sequence (derived from one of our random-sequence controls), with the 12-nt window tiled at 8-nt intervals across each of the three 50mers. Mutant 50mer con-

structs were generated by oligo annealing and primer extension, as described above for endogenous 50mer sequences.

U-rich sequence deletions (Fig. 2.6E-F): Candidate U-rich elements were deleted from the full-length *Hmga2* (or *let-7* disrupted *Hmga2*) reporter using overlap extension PCR; the resulting amplicons were inserted into pmirGLO, as described above. 200mer reporters containing the same mutations were PCR amplified from the mutated full-length *Hmga2* reporters and cloned as described above.

The inserts and cloning junctions of all mutated constructs were fully sequence-validated.

Dual-Luciferase reporter assays

Cells were seeded 24 h pre-transfection at densities appropriate for each cell line (70,000 cells/well for A549, 52,500 cells/well for HeLa, 50,000 cells/well for F9 and 3T3 cells) in 24-well plates. A549 and HeLa cells were transfected with Lipofectamine 2000 (Invitrogen); F9 and 3T3 cells were transfected with Lipofectamine LTX and Plus reagent (Invitrogen), following the manufacturer's protocols, and using 2.5-50 ng of pmirGLO-derived reporter plasmids, with the amount determined by the identity of the cell lines, to account for differences in transfection efficiency, and/or inherent differences in expression from the plasmid. Carrier DNA (pUC19) was included to increase transfection efficiency (100-140 ng/well). Cells were harvested 30 h post-transfection by removing the media, washing once with 1 PBS, and frozen at -80°C. Luciferase assays were performed using the Promega Dual-Luciferase Reporter Assay kit and a Veritas Microplate Luminometer (Turner Biosystems) according to manufacturer's protocols. The resulting Firefly values were first normalized to the Renilla values for each individual well, thus controlling for transfection efficiency. Normalized Firefly values were then scaled relative to the geo-

metric mean of normalized firefly values for the appropriate baseline control (i.e., the random-sequence controls for most experiments). The resulting data was plotted as log2 values. The error bars were estimated as the nonparametric equivalent to one standard deviation (~68% of the data is within the error bars). To determine significant differences between different reporters, a Wilcoxon rank-sum two-sample test was used. Multiple comparison corrections (Bonferroni) were used when appropriate (as indicated).

Quantifying sequence divergence and regulatory impact divergence between mouse, human, and chicken *Hmga2* 3′ UTRs

Sequence divergence was calculated as previously described [183], counting gaps as sequence differences. Divergence in regulatory impact was calculated as the absolute value of the deviation from the mean regulatory impact (across species), summed across all three species.

Calculating expected regulatory potential

We used a Monte Carlo sampling strategy to estimate an expected regulatory impact for a 3′-UTR fragment from reporter values corresponding to two constituent fragments (Figs. 2.3, 2.4). We used all individual normalized luciferase values (as described above) for each of the two constituent fragments and randomly sampled one from each distribution and multiplied them. This procedure was repeated 100,000 times, generating a distribution of simulated values, which represent the expected values assuming no regulatory interactions between the two constituent fragments (i.e., the model outlined in Fig. 2.3A). This distribution of simulated values was then treated equivalently to the distributions of all other reporter data (as described above).

Identifying significant outliers from observed-modeled regressions

We used a Monte Carlo sampling strategy to identify significant deviations in expected regulatory impact (y-axis values in Fig. 2.3). We simulated individual y-axis values by assuming that the true value lied precisely on the regression line, and then used the experimentally determined variations in reporter data to generate the simulated reporter data for that y-axis value. This procedure was repeated 100,000 times for each reporter fragment, generating a distribution of simulated values, which were used to estimate the probability that the actual value was a significant outlier. This same approach was also used to estimate the maximum possible correlations between observed and expected values for the data shown in Fig. 2.3.

shRNA knockdown experiments

shRNA hairpin plasmids (The RNAi Consortium, TRC) were used to generate shRNA virus according to TRC protocols [184]. Media supernatant with virus was harvested on day 2 and 3 post transfection and pooled.

A549 cells were plated at 80,000 cells/mL into 6-well plates (three wells per infection) and infected with shRNA virus 24 h later in polybrene media (DMEM 10% FBS, 8 g/mL polybrene [Sigma]) at an MOI of 4. Cells were selected with puromycin media (DMEM 10% FBS and 3 g/mL puromycin [Sigma]) and expanded to 75 cm^3 flasks. On day 4 post infection, cells were plated for transfection and luciferase assays (as above). Cells were also lysed in TRIzol for RNA isolation, following the manufacturer's recommended protocol (Life Technologies).

qRT-PCR

RNA isolated from shRNA knock down cells was treated with recombinant DNase I (Roche) for 20 min, and then phenol chloroform extracted. cDNA was generated using Thermo Scientific RevertAid Reverse transcriptase and an oligo(dT)18 primer at 42°C for 60 min; the reaction was heat inactivated at 70°C for 10 min. qPCR reactions were performed using Taq polymerase and SYBR Green (Life Technologies) as the detection agent and using GAPDH as a house-keeping gene to which to normalize. Each qPCR reaction was done in triplicate, and performed on at least two biological replicate samples.

CHAPTER 3

AN NMD-INDEPENDENT ROLE FOR 3′UTR LENGTH IN POST-TRANSCRIPTIONAL GENE REGULATION

[1]

3.1 Abstract

Though the role of 3′ UTR length in triggering the nonsense-mediated decay (NMD) pathway has been well established, its wider role in gene regulation remains to be determined. Using random-sequence 3′ UTR mimics of varying lengths, we have identified a novel role for 3′ UTR length in triggering an NMD-independent decay pathway in human cell lines. Reporter transcripts with random 3′ UTRs as short as 400 nucleotides (nts) were repressed by this pathway, with the repression growing stronger with increasing 3′ UTR length before tapering off at around 1600-2000nts. Though the exact mechanism of this novel pathway remains to be elucidated, we have determined that it affects the decay rate of mature mRNAs in a deadenylation-independent manner. In addition to random 3′ UTR mimics, 3′ UTRs generated from endogenous sequences were also destabilized by increasing 3′ UTR length, though the effect was weaker. Overall, our discovery demonstrates that 3′ UTR length can affect gene expression, independently of any regulatory sequence motifs, through mechanisms other than NMD.

[1]Though the experiments in this chapter were designed by the author, Katla Kristjánsdóttir, many (Fig. 3.1C; Fig. 3.3D,E; Fig. 3.4A,D,E,F) were executed in collaboration with Elizabeth Fogarty. In addition, experiments in Fig. 3.1D and E were performed by undergraduate students, Akila Ventakataramani and Ciara McDermott, working under the author's guidance.

3.2 Introduction

Post-transcriptional regulation adjusts the expression of a majority of mammalian transcripts, with important consequences for cellular function and disease. It is most often mediated by *trans*-acting factors, such as microRNAs (miRNAs) or RNA binding proteins (RBPs), which bind to regulatory sequence motifs within the 5′ or 3′ untranslated regions (UTRs) of the mRNA. Though both the 5′ and 3′ UTRs contain important regulatory information, the majority is contained within the 3′ UTR, partly because its length allows more space for regulatory motifs [5]. Post-transcriptional regulation allows for more varied and complex regulatory strategies than possible using transcriptional regulation alone, contributing significantly to the tissue-specificity of gene expression [185]. Therefore, it is not surprising that 3′ UTR length increases with the increasing morphological complexity of animals [186]. In addition to increasing the number of regulatory sequence motifs within a 3′ UTR, lengthening a 3′ UTR can, in and of itself, also have regulatory consequences [149, 144, 145, 147, 146, 143], though the extent to which 3′ UTR length contributes to gene expression remains unclear.

Several studies have found a negative correlation between 3′ UTR length and transcript stability, though the effect was not consistent between different cell types [149, 187, 143] (Yang et al. 2003, Sharova et al. 2009, Spies et al. 2013). However, interpreting these results is complicated, as the underlying cause of the correlation is not clear. The correlation may partially be explained by an increase in repressive *cis*-regulatory sequences in longer 3′ UTRs. However, as was demonstrated in chapter 2, and other studies (Wissink et al. 2016, Spies et al. 2013), activating regulatory sequences are at least as common as repressive ones. Long 3′ UTRs are, therefore, just as likely to have additional activating sequences as repressive ones. Beyond regulatory sequences, the best-known regulatory pathway for which 3′ UTR length is a factor is the nonsense mediated decay (NMD) pathway [188]. It is likely that this pathway at least partially explains the correlation between

3′ UTR length and expression levels. Though repressive regulatory elements and NMD are both likely to contribute to the observed repression of long 3′ UTRs, the existence of other regulatory pathways that respond to 3′ UTR length has not been ruled out. Therefore, it remains to be seen what proportion is explained by negative regulatory motifs, what proportion is explained by NMD, and whether there are any other possible causes.

Currently, the only known gene regulatory mechanisms that is thought to respond to 3′ UTR length independently of specific regulatory sequences is NMD. NMD is a surveillance mechanism that mostly targets and removes mutated or incorrectly processed transcripts with premature stop codons (PTCs) . However, NMD has also been shown to target and regulate normally processed, non-aberrant mRNAs [189, 188, 190]. Non-aberrant NMD targets generally contain features that, even when normally processed, indicate that translation was prematurely terminated. For example, transcripts with upstream open reading frames (uORFs) within the 5′ UTR or introns in their 3′ UTRs contain exon-exon junctions downstream of a terminating ribosome, which is a classical trigger for NMD. For transcripts with long 3′ UTRs, it is the unusually long distance between the stop codon and proteins at the polyA-tail that inhibit NMD that causes the transcripts to be recognized as NMD targets [189, 190].

In most simple organisms, such as in baker's yeast, 3′ UTRs are generally short, so a long 3′ UTR most often indicates aberrant processing [191, 192]. On the contrary, complex organisms such as mammals, and particularly humans, have evolved a large number of 3′ UTRs that exceed the length necessary to elicit NMD. In humans, the average length for 3′ UTRs is around 1300nts [5], whereas reporters with artificial 3′ UTRs as short as ~420 nts have been shown to elicit NMD [148, 145]. However, only a subset of long 3′ UTRs are NMD targets, with many escaping targeting through mechanisms that are only just beginning to be elucidated [193, 138, 139]. Thus, NMD functions both as a surveillance mechanism and as a general regulator of gene expression. The identification of NMD as

a regulatory mechanism that measures the 3′ UTR and regulates transcripts based on 3′ UTR length opens the door to the possibility that other regulatory mechanisms may do the same.

In this study, we find evidence of a strong negative relationship between 3′ UTR length and mRNA stability. Importantly, the relationship is almost entirely independent of NMD, indicating a novel role for 3′ UTR length in mRNA regulation. We observe this effect using both artificial and endogenous 3′ UTRs; however, the effect is weaker with endogenous 3′ UTRs. Thus, endogenous 3′ UTRs can, to some extent, avoid NMD-independent mRNA-destabilization caused by longer 3′ UTRs. This result parallels the previously published observation that long 3′ UTR transcripts escape NMD surveillance. Overall, this study re-emphasizes the importance of 3′ UTR length in regulating gene expression and identifies a novel, NMD-independent, decay mechanism influenced by 3′ UTR length.

3.3 Results

3.3.1 Random 3′ UTR-mimic reporters are strongly affected by 3′ UTR length

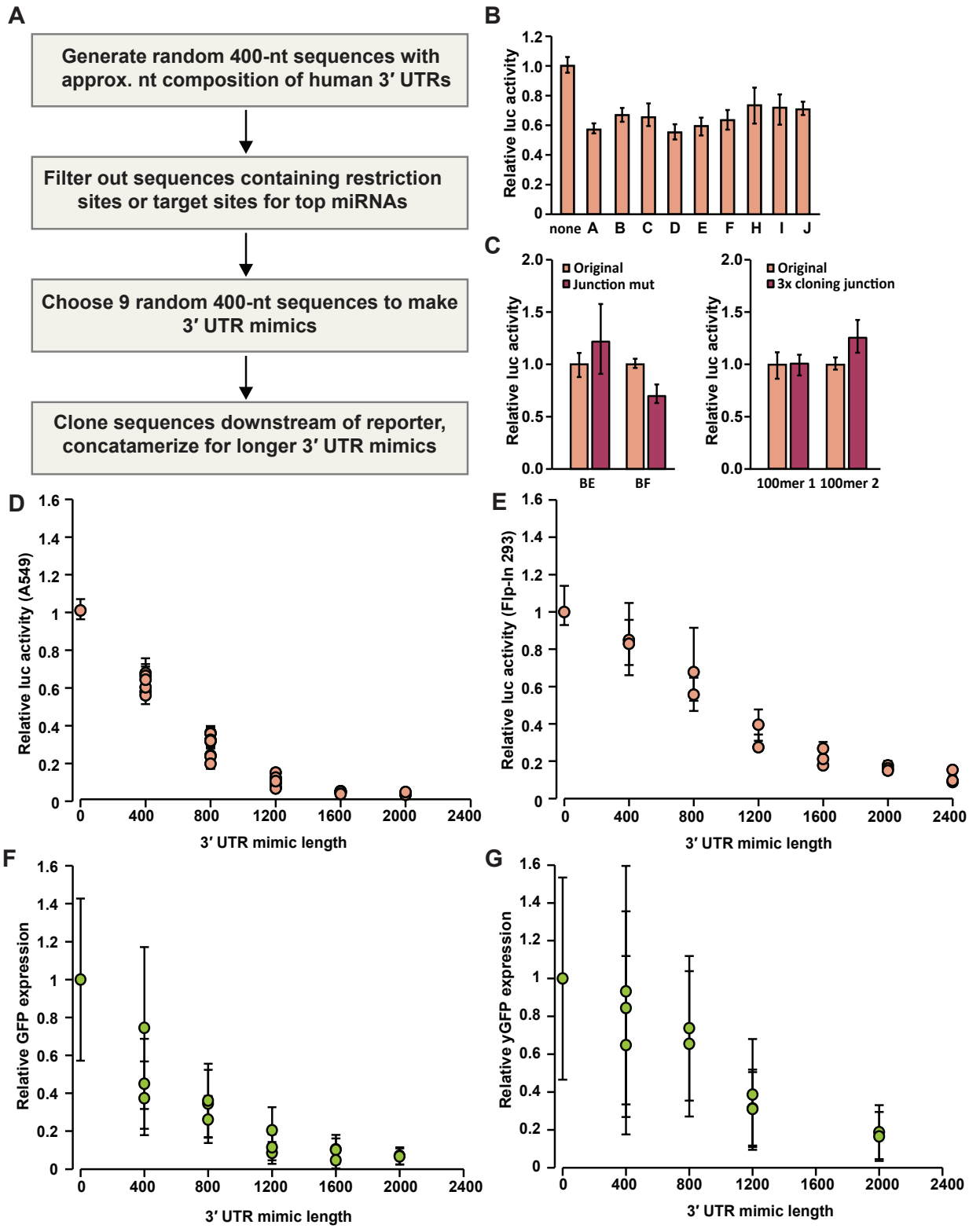
Previously, we studied the regulatory sequences within the *Hmga2* and *PIM1* 3′ UTRs (chapter 2, [125]). We did this by measuring the effect on reporter expression of small fragments of the 3′ UTRs, tiled across the length of each 3′ UTR. Based on the effect of the individual fragments on reporter expression, the overall effect of both 3′ UTRs should have been to enhance expression. However, when we measured the effect of the full-length 3′ UTRs, their overall effects were strongly repressive. We hypothesized that this

difference could, at least in part, be explained by the increase in 3′ UTR length. In support of this hypothesis, we also noted that control reporters with random, artificial, 3′ UTRs (50 to 400 nts in length) showed a small but steady decrease in expression with increasing 3′ UTR length. As the random controls can be assumed to contain no regulatory sequence elements, this effect is most likely due to the length of the transcript. In light of these results, we wanted to explore further the role of 3′ UTR length in gene expression. In particular, we were interested in systematically quantifying the effect of 3′ UTR length and determining to what extent the effect is attributable to NMD.

Distinguishing the effect of 3′ UTR length from the effect of regulatory sequences is a complicating factor in many studies that examine the role of 3′ UTR length on gene expression. To address this concern, our strategy was to use randomly generated sequences, similar to the controls used in chapter 2, of increasing length to mimic 3′ UTRs. As before, we reasoned that randomly generated sequences were unlikely to contain functional regulatory motifs. To control for any possible effect of nucleotide composition, the likelihood of each base at any position was set to be proportional to their overall occurrence rate in human 3′ UTRs. In addition, to further reduce the number of regulatory sequences within our random 3′ UTR mimics, we excluded any sequences containing target sites for the top 20 miRNAs in the cell lines to be used in the study [9]. We call these random sequences "3′ UTR mimics" (Fig. 3.1A).

We wanted to systematically examine the effect of 3′ UTR length across a wide spectrum of lengths. To do this we use 400-nt 3′ UTR mimics (400mers) as the building blocks to make ever larger random 3′ UTR mimics. We chose nine random 400mers and cloned them downstream of a luciferase reporter. To ascertain that our 400mer 3′ UTR mimics contain no strong regulatory sequences, we measured their effect in luciferase reporter assays. All nine of the 3′ UTR mimics had almost identical effects on reporter expression, strongly indicating they do not contain regulatory sequence information (Fig. 3.1B). We

Figure 3.1: 3′ UTR mimics reveal a strong negative effect of 3 ′ UTR size on reporter expression.
The design process for the 3′UTR mimics is illustrated in (A). The reporters used in this study are Firefly luciferase and two GFP reporters, GFP and yGFP, which have varying stabilities (due to codon optimization for mammal and yeast expression, respectively). Random 3′ UTR mimics were concatamerized to make mimics up to 3600-nt long. (B) The relative luciferase activity for all 9 random 400mer 3′ UTR mimics, relative to a no-insert control, in transient transfections in A549 cells ($n \geq 9$, error bars show non-parametric estimate of one standard deviation (NP-SD)—see Methods). (C) Left: The effect of mutating the 6-nt residual cloning sequence between concatamerized random 400mers in 800mer 3′ UTR mimics BE and BF. Right: The effect of adding three copies of the 6-nt residual cloning sequence to random 100mer 3′ UTR mimics [125]. For both Left and Right, bars represent relative luciferase expression after transient transfection into A549 cells normalized to the original unaltered construct ($n = 9$, error bars are NP-SD). (D)-(G) Relative reporter expression of progressively longer 3′ UTR mimics (see Table 3.1) after transient transfections of luciferase reporters into A549 (D) or FlpIn293 (E) cells, measured by luciferase activity ($n \geq 9$, error bars are NP-SD), or after stable integration of GFP (F) or yGFP (G) reporters into a unique genomic locus in FlpIn293 cells, measured by GFP fluorescence (10,000 cells measured, error bars represent one standard deviation (SD)). Each circle represents an individual, unique, 3′ UTR mimic.


A
Generate random 400-nt sequences with approx. nt composition of human 3′ UTRs
Filter out sequences containing restriction sites or target sites for top miRNAs
Choose 9 random 400-nt sequences to make 3′ UTR mimics
Clone sequences downstream of reporter, concatamerize for longer 3′ UTR mimics
B
Relative luc activity
none A B C D E F H I J
C
Original
Junction mut
BE BF
Original
3x cloning junction
100mer 1 100mer 2
D
Relative luc activity (A549)
3′ UTR mimic length
E
Relative luc activity (Flp-In 293)
3′ UTR mimic length
F
Relative GFP expression
3′ UTR mimic length
G
Relative yGFP expression
3′ UTR mimic length

then used a sequential cloning strategy to concatenate the 400mers into various 800 to 3200-nt long 3′ UTR mimics, covering the majority of the human 3′ UTR length distribution (see Table 3.1).

Table 3.1: Random 400mer 3′ UTR mimics and concatenated longer 3′ UTR mimics. Stars indicate 3′ UTR mimics used for integrated reporters.

400mers	800mers	1200mers	1600mers	2000mers	2400mers	2800mers	3200mers
A	BC	BEC*	BEFC	BEFAH	BEFAHD	BEFAHDI	CDBFEAHI
B*	BD	BED	BEFA	CDBEF*	CDBEFI	CDBEFIH	
C*	BE	BEF	BEFD*	CDBFE*	CDBFEA	CDBFEAH	
D	BF*	CFB	CDBE				
E*	CB	CFD	CDBF*				
F	CD*	CFE	CFDB				
H	CE*	CDB	CFDE*				
I	CF	CDE*					
J		CDF*					

To determine the effect of 3′ UTR length, each 3′ UTR mimic listed in table 3.1 was cloned into a luciferase reporter and assayed in A549 cells (Fig. 3.1D). We see a dramatic negative response of reporter expression to increasing 3′ UTR length, strongly indicating that long 3′ UTRs repress expression independently of any regulatory sequence motifs. The expression of the 3′ UTR mimic reporters drops rapidly with increasing 3′ UTR length and levels off at 2-3% of the expression of a no-insert control reporter at around 1600 nts. However, as the strategy we employed to concatenate the 400mers leaves a 6-nt motif, a cloning junction, between each 400mer, it was possible that this cloning junction was responsible for the increasingly repressive effect of the 3′ UTR mimics. However, mutating the cloning junction in an 800mer 3′ UTR mimic or adding multiple copies of it to a 100mer 3′ UTR mimic has either no effect or an opposite effect to that expected if it were a repressive element (Fig. 3.1C).

To determine if the 3′ UTR-length effect observed in Fig. 3.1D was cell type specific, we repeated the experiment for a subset of the 3′ UTR mimic luciferase reporters, covering

the entire length-range, in Flp-In T-REx 293 cells (Table 3.1, Fig. 3.1E). In Flp-In T-REx 293 cells, reporter expression drops somewhat slower in response to 3′ UTR length than in A549 cells and tapers off later (at ~2000 nts) with final expression level at 10% of no-insert control expression. This difference between A549 cells and Flp-In T-REx 293 cells indicates there can be cell-type specific differences in the role 3′ UTR length plays in gene regulation. However, in both cell types, increasing 3′ UTR length had a potent effect on reporter expression. In both cases (Fig. 3.1D - E) we only show data up to the point where the length-effect tapers off. From these data, it seems there is a very strong effect of 3′ UTR length, at least in transiently transfected reporters, and that this effect varies somewhat depending on cellular context.

To exclude the possibility that the observed repressive effect of 3′ UTR length is an artifact of plasmid-derived expression, we generated GFP reporters and integrated them into a specific locus in the Flp-In T-REx 293 genome (Table 1). As these reporters contain an intron, expression and processing of the reporter transcript should mirror endogenous expression of typical mRNAs. These intron-containing integrated reporters also respond strongly to the length of the 3′ UTR mimic (Fig. 3.1F). In fact, expression closely follows the transient reporters in the same cell type and tapers off at 12% of no-insert control expression. Because of this similarity, integrated and transient reporters are used interchangeably in the remainder of this chapter. Additionally, as the GFP reporters use a different promoter than the luciferase reporters, we can exclude the possibility that the 3′ UTR length effect is a promoter specific phenomenon.

The GFP construct described above produces a particularly stable transcript (half-life >12 hours), which means it takes a long time for it to reach a new steady state after perturbation. To generate reporter cell lines that respond more quickly, we replaced the GFP open reading frame (ORF) in our 3′ UTR mimic reporters with one that produces a less stable GFP transcript, due to its codon usage being optimized for yeast. This resulting

in the inclusion of multiple codons that are non-optimal in mammals, which destabilizes the transcript. These yeast codon-optimized GFP (yGFP) reporters were integrated into Flp-In T-REx 293 cells as described above. As expected, already destabilized yGFP reporters respond less drastically to 3′ UTR mimic length (Fig. 3.1G). However, the effect is still quite strong, with 2000mer 3′ UTR mimic reporters expressed at ~20% of no-insert control expression.

To summarize, we successfully created 3′ UTR mimics of varying lengths that contain no apparent strong regulatory sequences in the cell lines we used. Reporters containing these 3′ UTR mimics display a dramatic negative response to increasing 3′ UTR mimic length. This response is independent of the mode of transcription (promoter identity and plasmid or genomic transcription) and the inclusion of an intron.

3.3.2 Repressive effect of 3′ UTR mimic length is largely NMD-independent

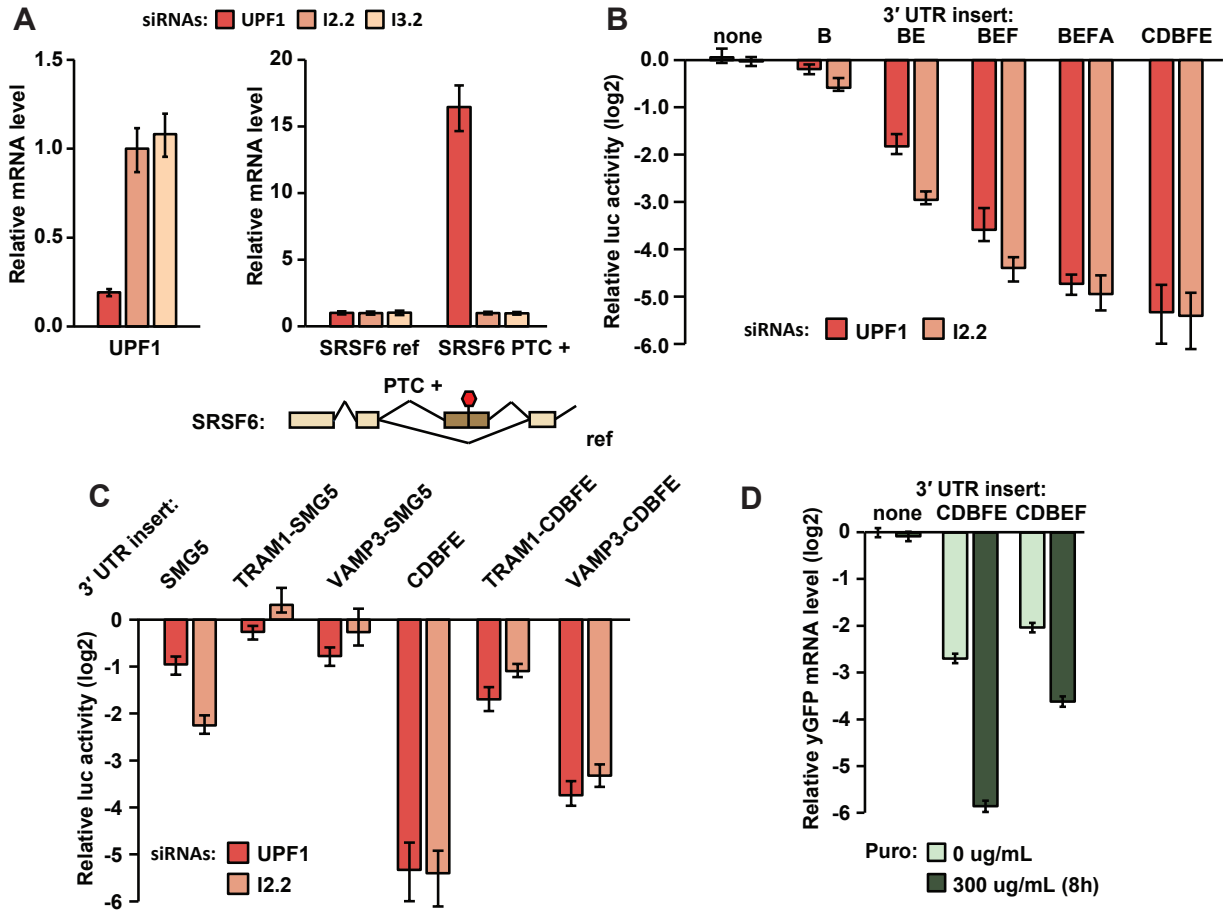
It has already been established that nonsense-mediated decay can be triggered by long 3′ UTRs, and that artificial 3′ UTRs are particularly sensitive to this mode of NMD activation [147, 146, 138, 148, 145, 193]. To determine to what extent our 3′ UTR mimics are affected by NMD we knocked down UPF1, a critical component of the NMD pathway, using a previously described siRNA in HeLa cells [194]. As controls, we used inert siRNAs (I2.2 and I3.2) specifically designed in our lab to be non-targeting (due to weak binding affinity and large number of targets). We confirmed decreased UPF1 transcript levels after knockdown (KD) using qRT-PCR (Fig. 3.2A, left). As a measure of NMD disruption, we also measured the increase in the transcript level of an endogenous target of NMD, SRSF6 (Fig. 3.2A, right). Alternative splicing creates either a reference isoform that is not an NMD target, or an isoform that contains a premature stop codon (PTC) and is therefore

a classical target of NMD [195]. Upon UPF1 KD, the non-targeted isoform is unaffected but the PTC isoform is strongly stabilized, illustrating that NMD has been successfully disrupted.

Upon UPF1 KD, there is very little change in the 3′ UTR-length mediated repression of 3′ UTR mimic reporters, indicating that NMD does not have a strong effect (Fig 3.2B). Repression is somewhat diminished by UPF1 KD in the shorter constructs, but this effect quickly weakens as the 3′ UTR mimics get longer. The fact that NMD-sensitivity is inversely related to 3′ UTR mimic length could mean one of two things: One, longer 3′ UTR mimics have such a strong affinity for UPF1 that they are insensitive to this level of UPF1 KD. Two, 3′ UTR mimics are weakly targeted by NMD but this effect is crowded out in the larger constructs by another, much stronger, 3′ UTR-length dependent repressive mechanism.

To attempt to distinguish which model better explains the weak effect of UPF1 KD on 3′ UTR mimic reporter expression, we took orthogonal approaches to determine if our long 3′ UTR mimics are indeed targets of NMD. Our first approach took advantage of a recent study that demonstrated that the first 200 nts of long endogenous 3′ UTRs that escape NMD can protect otherwise NMD-targeted 3′ UTRs when added just downstream of the stop codon [138]. We isolated two such regions, from the human TRAM1 and VAMP3 3′ UTRs, and cloned them upstream of either the SMG5 3′ UTR (a 3′ UTR known to elicit NMD) or a 2000mer 3′ UTR mimic in luciferase reporters (Fig. 3.2C). For both 3′ UTRs, the addition of these NMD-escape regions lead to diminished repression of reporter expression, though the effect was much weaker in constructs with the VAMP3-derived region. However, when combined with UPF1 knockdown, there is hardly any difference between the SMG5 3′ UTR reporters with and without the NMD-escape regions, indicating the NMD-escape regions provide no additional protection from NMD when UPF1 is knocked down. This strongly suggests UPF1 and NMD-escape regions

Figure 3.2: 3′ UTR-length dependent repression observed for 3′ UTR mimic reporters is largely NMD independent.
(A) NMD disruption by UPF1 knockdown indicated by the change in transcript level upon 72 hour siRNA depletion of UPF1 in HeLa cells, compared to inert siRNAs (I2.2 and I3.2) as measured by qRT-PCR (n = 6, two biological replicates, each in triplicate, error bars indicate one SD). Left: change in UPF1 transcript levels. Right, top: Change in transcript levels for two isoforms of the SRSF6 transcript (depicted below), one who is not an NMD target (ref) and one who is a strong NMD target (PTC+). Right, bottom: Model showing the difference between the two SRSF6 transcripts. Gray boxes represent exons and lines represent splicing events. Alternative exon in darker grey with premature termination codon (PTC) marked by a red hexagon. Splicing isoforms measured above are indicated by ref and PTC+. (B) The effect of UPF1 knockdown (performed as in (A), except using only I2.2 as control), on expression of transiently transfected luciferase reporters with different length random 3′ UTR mimics (B, BE, BEF, BEFA and CDBFE), normalized to a no-insert control (n = 12, three independent siRNA transfections, error bars are one NP-SD). (C) The effect of UPF1 knockdown (performed as in (B)), on expression of transiently transfected luciferase reporters with different 3′ UTRs: the SMG5 3′ UTR and the 2000mer random 3′ UTR mimic CDBFE, with and without the addition of the first 200nt of the TRAM1 or the VAMP3 3′ UTRs upstream of them. Reporters are normalized to a no-insert control (not shown, n = 9, two independent siRNA transfections, error bars are one NP-SD). (D) The effect of translation inhibition by an 8 hour incubation with 300 µg/mL Puromycin on the transcript levels of integrated yGFP reporters in FlpIn293 cells containing 2000mer random 3′ UTR mimics (CDBFE and CDBEF), compared to a no-insert control. Transcript levels are normalized to the no puromycin no-insert control (n = 6, two biological replicates, each in technical triplicate, error bars are one SD).


A
siRNAs: UPF1 I2.2 I3.2
Relative mRNA level
1.5
1.0
0.5
0.0
UPF1
Relative mRNA level
20
15
10
5
0
SRSF6 ref
SRSF6 PTC +
PTC +
SRSF6:
ref
B
3′ UTR insert:
none
B
BE
BEF
BEFA
CDBFE
Relative luc activity (log2)
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
siRNAs: UPF1 I2.2
C
3′ UTR insert:
SMG5
TRAM1-SMG5
VAMP3-SMG5
CDBFE
TRAM1-CDBFE
VAMP3-CDBFE
Relative luc activity (log2)
0
-1
-2
-3
-4
-5
-6
siRNAs: UPF1
I2.2
D
3′ UTR insert:
none
CDBFE
CDBEF
Relative yGFP mRNA level (log2)
0
-1
-2
-3
-4
-5
-6
Puro: 0 ug/mL
300 ug/mL (8h)

are redundant and are working through the same pathway. For the long 3′ UTR mimic, however, there is no effect of UPF1 knockdown on any of the reporters. This means that the NMD-escape regions still diminish the repression of the reporter, even when UPF1 is knocked down. This could be due to the fact that the UPF1 KD is not strong enough to disrupt NMD for this particular target, while the addition of the NMD-escape regions is. However, given that we detect robust knock down of UPF1 mRNA, it is more likely that the long 3′ UTR mimic is not a strong target of NMD and that the NMD escape regions can also protect transcripts from other forms of decay.

A final approach to determine the extent of NMD at long 3′ UTR-mimic transcripts, relies on the fact that a physical interaction between UPF1 and the terminating ribosome is necessary to trigger NMD [188]. A consequence of this is that NMD cannot affect transcripts unless they have been translated and contain a terminating ribosome. Puromycin inhibits translation by inducing premature chain termination, leading to ribosome release [196]. Therefore, if NMD is the primary mechanism by which our long 3′ UTR mimics lead to transcript repression the repressive effect should be diminished or completely abrogated by treatment with puromycin [194, 197]. We treated cells expressing yGFP constructs containing a no-insert control or two different 2000mer 3′ UTR mimics with 300 µg/mL puromycin for 8 hours. We then measured yGFP mRNA levels using qRT-PCR and saw that not only did the 3′ UTR-length mediated repression not diminish, but was amplified by a large margin (Fig. 3.2D). It seems, therefore, that our long 3′ UTR mimics are not targets of NMD, but of a different mechanism that may be repressed, rather than activated, by translation.

3.3.3 Deadenylation-independent decay of mature 3′ UTR-mimic transcripts

After establishing that NMD is unlikely to be the primary mechanism through which our long 3′ UTR-mimic reporters are being repressed, we worked to narrow down the options for possible alternative mechanisms. As we saw in Fig. 3.2D, the repression of long 3′ UTR mimic reporters is measurable at the mRNA level and the repression is similar to what we saw for protein levels in the same cell lines (see Fig. 3.1G). This indicates that the repression acts to reduce the level of mRNA, rather than decreasing translation. In fact, when we plot relative GFP protein levels (measured by flow cytometry) as a function of the corresponding mRNA levels (measured by qRT-PCR) for constructs with varying 3′ UTR lengths, the two correlate almost perfectly with a slope of 0.97 (Fig. 3.3A). This means that the entirety of the repression observed can be explained by a reduction in mRNA levels, and not by any regulatory mechanisms that affect translation.

There are two primary ways to change the amount of mRNA present in a cell at steady state: changing its synthesis rate (transcription or processing) or its decay rate. A simple way to measure whether changes in transcription or mRNA processing are involved is to measure the effect of 3′ UTR length on pre-mRNA levels. If the amount of pre-mRNA decreases with increasing 3′ UTR length, this would indicate a decrease in transcription (or, less likely, an increase in splicing efficiency). Conversely, if the amount of pre-mRNA increases with increasing length, this would indicate a defect in splicing (or, less likely, an increase in transcription). We used qRT-PCR primers that specifically bind unspliced yGFP reporter transcripts to measure pre-mRNA levels in yGFP reporter cell lines with different length 3′ UTR mimics (Fig. 3.3B). Though there is variation in the measured pre-mRNA levels between the different reporters, we found no relationship between pre-mRNA levels and 3′ UTR length. This indicates that the repressive effect of 3′ UTR mimic

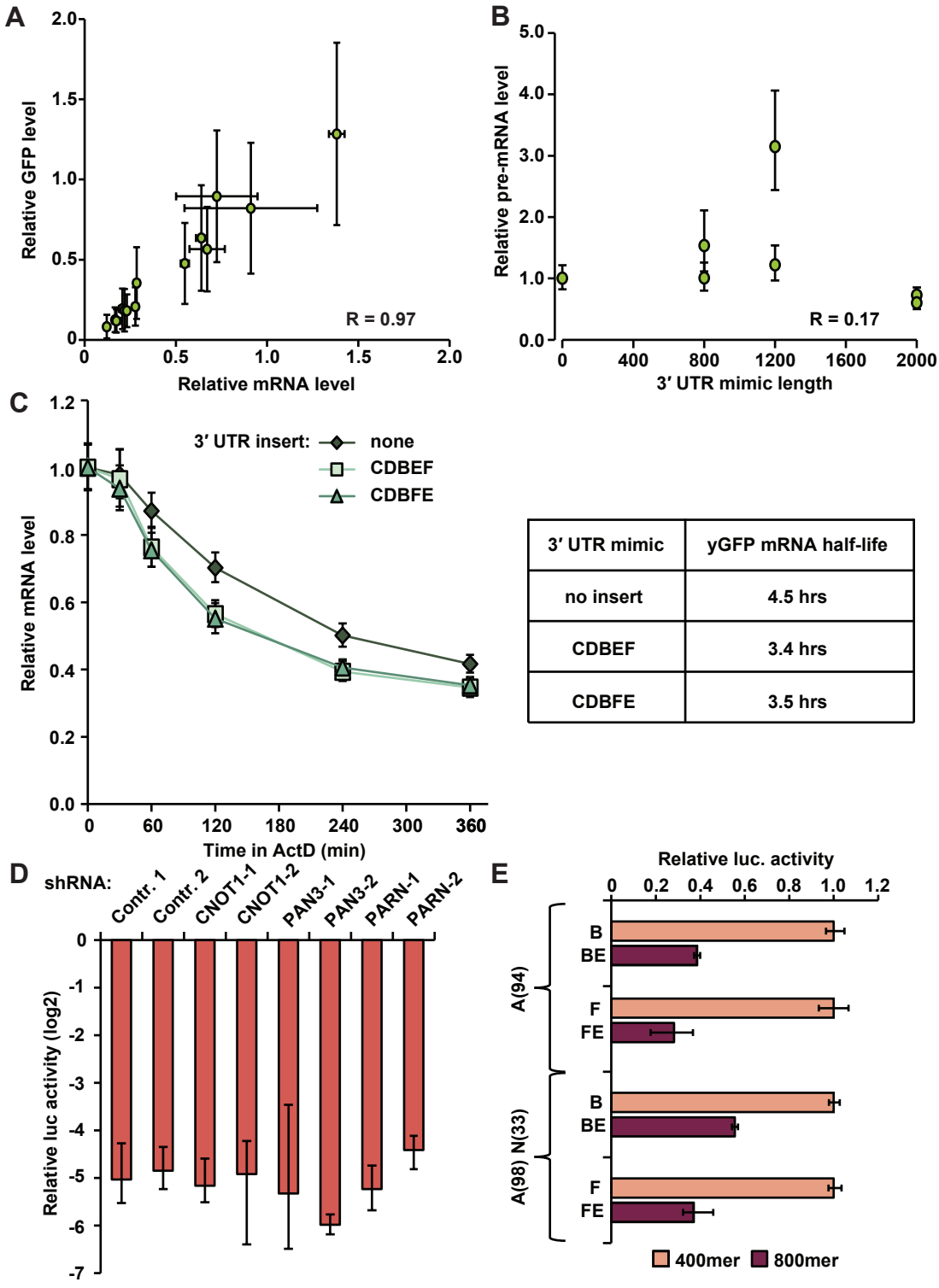
length is independent of transcription and splicing. This was a likely conclusion, given the similarity in the 3′ UTR length effect between the two reporter types, which are expressed from different promoters and only one of which contains an intron.

When mRNA synthesis has been eliminated as the source of 3′ UTR-length determined differential expression, the most likely mechanism that remains is differential mRNA decay. To determine whether decay is the primary mechanism by which longer 3′ UTRs lead to reporter repression, we measured the half-life of integrated yGFP reporter transcripts with either a no-insert control 3′ UTR or one of two different 2000mer 3′ UTR mimics. We did this by inhibiting transcription with actinomycin D and monitoring the change in reporter transcript levels at different time points (Fig. 3.3C). We found that the long 3′ UTR-mimic reporters did indeed have shorter half-lives than a control reporter that has no inserted 3 UTR sequence, confirming that differential expression between the short and long 3′ UTR mimics is due to differential decay.

The most common rate-limiting step in mRNA decay is deadenylation of the polyA tail and mammalian cells contain many different deadenylase complexes [198]. To determine if deadenylation is necessary for the increased decay rate observed for long 3′ UTR-mimic reporters, we used RNAi to knock down essential components of three major deadenylase complexes and measured the effect on a 2000mer 3′ UTR-mimic luciferase reporter, compared to a no-insert control (Fig. 3.3D). Though knockdown was successful for all deadenylase components targeted (data not shown), none showed a consistent effect on the expression of the reporter. While this result suggests that none of these complexes are involved, it does not exclude the possibility that other, less common, deadenylase complexes may be involved.

To more directly measure if deadenylation has a role in the observed 3′ UTR-length-mediated decay, we *in vitro* transcribed luciferase reporter mRNAs with either a regular polyA tail or a tail protected by a downstream non-polyA sequence and measured lu-

Figure 3.3: Deadenylation-independent decay of mature mRNA in the cytoplasm is the cause of differential expression of 3′ UTR-mimic reporters.
(A) The relationship between reporter protein and mRNA levels are illustrated by plotting the relative GFP intensity (10,000 cells measured, error bars represent one SD), of integrated 3′ UTR-mimic reporters ranging in size from 400 to 2000 nts (see Table 3.1), as a function of the relative mRNA levels for the same cell reporters, measured by qRT-PCR ($n = 3$ technical replicates, error bars represent one SD). In both cases, reporter levels were normalized to the average of three random 400mer reportes. The correlation has a Pearson correlation coefficient $R = 0.98$ and $P < 1 \times 10^{-5}$. (B) The relationship between pre-mRNA levels and 3′ UTR length is illustrated by plotting the relative levels of integrated yGFP pre-mRNA reporters with 3′ UTR mimics of varying length, normalized to a no-insert control, as a function of 3′ UTR mimic length ($n = 6$, two biological replicates, each in technical triplicate, error bars represent one SD). The correlation has a Pearson correlation coefficient of 0.17 and $P > 0.05$. (C) Left: Decay kinetics of yGFP reporter mRNAs, with no insert or 2000mer 3′ UTR mimics CDBEF and CDBFE, analyzed by qRT-PCR after treatment with Actinomycin D ($n = 9$, 3 biological replicates, each in technical triplicate, error bars represent one SD). Right: Table summarizing mRNA half-lives calculated based on the data shown on Left. (D) The relative expression ($\log_2$) of a transiently transfected luciferase reporter with 2000mer 3UTR mimic CDBFE, normalized to no-insert control, in A549 cells after deadenylase disruption with shRNAs targeting key components ($n = 6$, two independent transductions for each shRNA, error bars represent one NP-SD). Each deadenylase component is targeted by two different shRNA hairpins (labeled 1 and 2) and two control hairpins are used which target GFP, which is not expressed in these cells. (E) The relative expression of 800mer 3′ UTR mimic luciferase reporters, normalized to 400mer-3′ UTR mimics, after transient transfection of *in vitro* transcribed mRNAs into A549 cells ($n = 9$, error bars represent one NP-SD). *In vitro* transcribed mRNAs contain either a regular polyA tail (A(94)) or a protected polyA tail (A(98)N(33)).



3′ UTR mimic	yGFP mRNA half-life
no insert	4.5 hrs
CDBEF	3.4 hrs
CDBFE	3.5 hrs

ciferase expression after transfection into A549 cells (Fig. 3.3E). As deadenylases have a much higher affinity for polyA sequences, the addition of the non-polyA sequence provides some protection from deadenylation. Therefore, if deadenylation is a crucial step in eliciting decay of transcripts with longer 3′ UTRs the addition of a protecting sequence should mitigate the repressive effect of increased 3′ UTR length. Because the length of the transcript may affect transfection efficiency, we limited the length difference by comparing 800mer 3′ UTR mimics with 400mer 3′ UTR mimics, a transcript length difference of only ~15%, which still lead to a ~40-50% decrease in reporter expression when using plasmids (see Fig. 3.1C). Our results show no real difference between reporters with unprotected and protected polyA tails, strongly indicating that deadenylation is irrelevant to the destabilization of longer 3′ UTR-mimic reporters. In addition, the fact that the different length 3′ UTR reporters still show differential expression, even when mature mRNA is transfected rather than processed endogenously, definitively excludes the involvement of nuclear events such as transcription or mRNA processing.

To summarize, we have determined that the repression of reporters with long 3′ UTR mimics is independent of transcription, mRNA processing and translation, and instead appears to be caused by increased mRNA decay. We have also established that the rate-limiting step for this decay mechanism is unlikely to be deadenylation. Further study is needed to determine exactly which decay pathway is being triggered by these long 3′ UTR mimics and to identify the *trans*-factors that are involved in regulating this decay pathway.

3.3.4 3′ UTR-length mediated repression of endogenously-derived 3′ UTRs

Though we attempted to mimic the qualities of endogenous 3′ UTRs when designing our 3′ UTR mimics, they are necessarily not perfect mimics as they are also designed to have minimal regulatory sequence information. Therefore, we cannot rule out that the repressive effect of 3′ UTR length we see using these mimics is caused by some unique property not shared by endogenous 3′ UTRs. We have evidence from our previous work that 3′ UTR length might strongly influence gene expression in real 3′ UTRs [125]. In studying the *Hmga2* 3′ UTR, we made multiple truncations of the 3′ UTR and measured the regulatory potential of these constructs in luciferase reporter assays. When the effect of each truncation on gene expression is plotted as a function of 3′ UTR fragment length, a striking negative correlation is observed (Fig. 3.4A). This correlation may be caused by a steady increase in negative regulatory motifs. However, this hypothesis is not supported by our previous analysis of the regulatory motifs contained within the *Hmga2* 3′ UTR [125].

As endogenous 3′ UTR sequences contain regulatory motifs, it is more difficult to distinguish the effect of increasing 3′ UTR length from a cumulative effect of an increasing number of negative regulatory motifs. One strategy to distinguish between these possibilities is to clone shorter endogenous 3′ UTR sequences downstream of reporters and then concatenate them into larger 3′ UTRs similar to how the random 3′ UTR mimics were constructed. We can then measure the regulatory effect of the individual, endogenously-derived 3′ UTR pieces and use that information to model an expected regulatory effect for the larger concatenated 3′ UTRs. In this model, the effect of a concatenated 3′ UTR is calculated as the product (log additive) of the regulatory impact of the smaller fragments (Fig. 3.4B). Because the model assumes there is no effect of 3′ UTR length the calculated

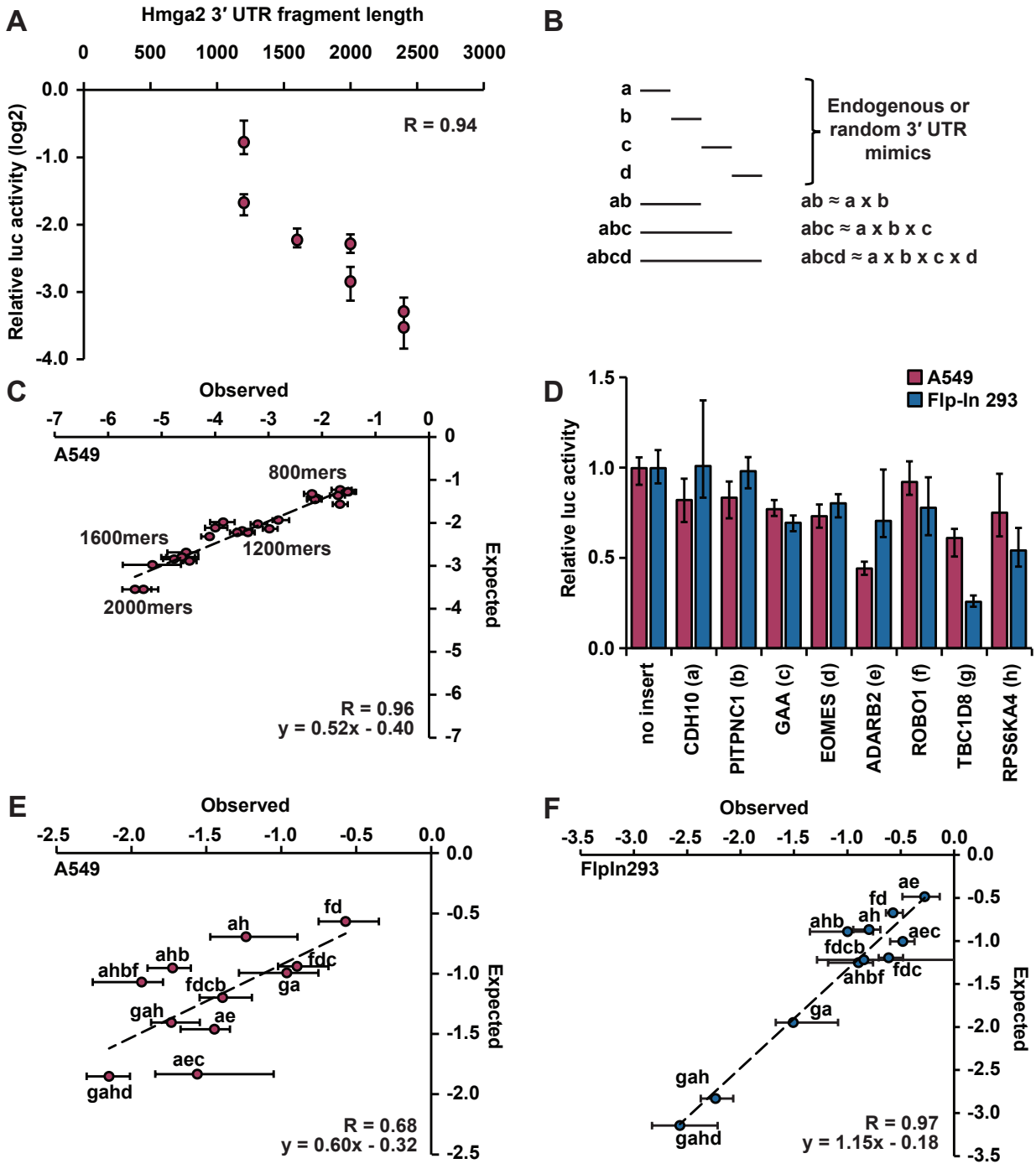
expectation will not take length into account. Therefore, if there is no length effect, a regression between the observed and expected regulatory impact of concatenated 3′ UTRs will have a slope of 1. In contrast, if there is a repressive length effect, the observed regulatory effect of concatenated 3′ UTRs will be more repressive than the calculated expectation, leading the slope of the regression to be less than 1. When we performed this analysis on our set of random 3′ UTR mimics, we saw that the observed and expected values correlated quite well ($R = 0.96$); however, in keeping with the demonstrated 3′ UTR length effect, the regression deviated from 1, with a bias towards lower observed expression (Pearson regression slope of 0.52, Fig. 3.4C).

Table 3.2: Fragments of human endogenous 3′ UTRs and concatenated longer 3′ UTRs

	Parent gene(s)	ID	size
1x	CDH10	a	362
	PITPNC1	b	521
	GAA	c	378
	EOMES	d	427
	ADARB2	e	586
	ROBO1	f	529
	TBC1D8	g	511
	RPS6KA4	h	537
2x	CDH10-ADARB2	ae	952
	CDH10-RPS6KA4	ah	900
	ROBO1-EOMES	fd	960
	TBC1D8-CDH10	ga	877
3x	CDH10-ADARB2-GAA	aec	1361
	CDH10-RPS6KA4-PITPNC1	ahb	1423
	ROBO1-EOMES-GAA	fdc	1369
	TBC1D8-CDH10-RPS6KA4	gah	1413
4x	CDH10-RPS6KA4-PITPNC1-ROBO1	ahbf	1954
	ROBO1-EOMES-GAA-PITPNC1	fdcb	1892
	TBC1D8-CDH10-RPS6KA4-EOMES	gahd	1842

To make the same comparison for endogenous 3′ UTR sequences we chose nine human 3′ UTR fragments previously isolated in our laboratory, each of which is between

Figure 3.4: Expression of reporters with 3 UTR mimics derived from endogenous 3 UTR sequences are also affected by 3 UTR size.
(A) The relationship between 3′ UTR size and relative luciferase reporter expression in A549 cells for different truncations of the *Hmga2* 3′ UTR [125], normalized to three random 400mer 3′ UTR mimics. The correlation has a Pearson correlation coefficient of $R = 0.94$. (B) An illustration of the regulatory impact of different 3′ UTR fragments when assuming no effect of 3′ UTR length or interactions between *cis*-elements. The cumulative regulatory effect of a 3′ UTR fragment is modeled as the product of the regulatory impacts of constituent smaller segments. (C) The observed regulatory impact of random 3′ UTR mimics of different sizes (x-axis) in A549 cells compared to the calculated expectation (y-axis) based on multiplying the effect of the constituent random 400mer 3′ UTR mimics, all normalized to a no-insert control ($n = 6$, error bars represent one NP-SD). The dashed line indicates the Pearson best-fit regression with $R = 0.96$ and a slope of 0.52. Reporters cluster based on 3′ UTR size, and the size is labeled next to each cluster. (D) The relative luciferase activity for 9 endogenously derived 3′ UTR fragments (see Table 3.2), relative to a no-insert control, in transient transfections in A549 or Flp-In T-REx 293 cells (FLp-In 293, $n \geq 6$, error bars are NP-SD). (E)-(F) As in (C), except 3′ UTRs are concatamerized endogenously derived 3′ UTR fragments (see Table 3.2) assayed in A549 cells (E) and Flp-In T-REx 293 cells (F) ($n = 6$, error bars represent one NP-SD). The Pearson R values are 0.67 and 0.97, respectively, and the slopes are 0.62 and 1.15, respectively.


A
Hmga2 3′ UTR fragment length
0
500
1000
1500
2000
2500
3000
Relative luc activity (log2)
0.0
-1.0
-2.0
-3.0
-4.0
R = 0.94
B
a
b
c
d
ab
abc
abcd
Endogenous or random 3′ UTR mimics
ab ≈ a x b
abc ≈ a x b x c
abcd ≈ a x b x c x d
C
Observed
A549
800mers
1600mers
1200mers
2000mers
Expected
R = 0.96
y = 0.52x - 0.40
D
A549
Flp-In 293
Relative luc activity
1.5
1.0
0.5
0.0
no insert
CDH10 (a)
PITPNC1 (b)
GAA (c)
EOMES (d)
ADARB2 (e)
ROBO1 (f)
TBC1D8 (g)
RPS6KA4 (h)
E
Observed
A549
fd
ah
ahb
fdc
ahbf
ga
fdcb
gah
ae
aec
gahd
Expected
R = 0.68
y = 0.60x - 0.32
F
Observed
FlpIn293
ae
fd
ah
ahb
aec
fdcb
fdc
ahbf
ga
gah
gahd
Expected
R = 0.97
y = 1.15x - 0.18

350-650 nts in length. The regulatory effect each of these has on reporter expression was determined (Fig. 3.4D) and then they were concatenated into larger chimeric 3′ UTRs (see Table 3.2 for full list of 3′ UTRs and concatamers). One concern when using endogenous 3′ UTR sequences is that regulatory elements within adjacent 3′ UTR fragments may not function independently of one another, as our model assumes. However, such interactions between elements are likely rare (see Chapter 2) and almost certainly equally likely to increase as they are to decrease expression. Therefore, these interactions, if any, would hopefully cancel each other out when examining the overall correlation. We plotted the calculated expected regulatory impact of the concatenated human 3′ UTRs as a function of their observed impact (Fig. 3.4E). As expected for endogenously-derived 3′ UTR sequences, the correlation between observed and expected effect of concatenated sequences is not as strong as for the random sequences; though it is still clearly significant ($R = 0.68$, $P < 0.05$). Despite a poorer correlation, there is still a clear bias towards lower observed expression, though it is weaker than for random 3′ UTR mimics (Pearson slope of 0.60). This indicates that the 3′ UTR-length-dependent repression we saw for our random 3′ UTR mimics also affects real 3′ UTR sequences. However, when we repeated this experiment in Flp-In T-REx 293 cells, the regression between the observed and expected values had a slope close to 1, indicating no effect of 3′ UTR length (Fig. 3.4F). This indicates there is a real difference in how our random 3′ UTR mimic reporters and our concatenated endogenous 3′ UTR reporters respond to 3′ UTR length, at least in certain contexts. Further study is needed to explore this difference.

3.4 Discussion

The extent to which 3′ UTR length contributes to gene regulation has long been unclear. Previous studies have demonstrated a genome-wide trend towards decreased stability for

transcripts with long 3′ UTRs in some cell lines [149, 8, 143], though not in others [187]. Until now, the destabilization of long 3′ UTRs has been attributed to an increased number of repressive sequence elements or to the triggering of NMD. The recognition of many long 3′ UTRs by the NMD surveillance machinery is well known, and the mechanisms through which other long 3′ UTRs escape recognition are beginning to be elucidated [138, 139]. However, it remained to be determined whether other regulatory pathways could be affected by 3′ UTR length, as well. In this study, we have used random 3′ UTR mimics to describe a second decay pathway that is triggered by 3′ UTR length and may be partially responsible for the global destabilization of transcripts with long 3′ UTRs.

The decay pathway we have identified does not require UPF1 nor the presence of a terminating ribosome, both of which are required for NMD, demonstrating that it is NMD-independent. This result was somewhat surprising, given that previous studies had shown that artificial 3′ UTRs are particularly likely to trigger NMD, presumably because they lack any mitigating factors that endogenous 3′ UTRs may have accumulated to avoid being targeted [148]. The fact that the shorter 3′ UTR mimics were affected by UPF1 KD, whereas the longer ones were not, indicates that the random 3′ UTR mimics can be targeted by NMD. We hypothesize, therefore, that NMD targeting is no longer effective if the target is already strongly repressed by other means, as is the case for the longer random 3′ UTR mimic reporters. This overshadowing, or repression, of NMD by other decay mechanisms has also been suggested for certain unstable transcripts with long 3′ UTRs in *Drosophila* and by the inhibition of NMD in mice by Ago2, a critical component of the miRNA-induced silencing complex (miRISC) [199, 200].

While the mechanism of this new 3′ UTR-length mediated decay pathway remains to be fully elucidated, we have shown that it affects the half-life of mature mRNAs, as *in vitro* transcribed reporter mRNAs were effectively repressed, and gets progressively stronger as the length of the 3′ UTR increases until it levels off at what we assume is

its maximal repressive capacity. We have also shown that this decay pathway does not require deadenylation, which leaves decapping or cleavage by endonucleases as likely mechanisms. Importantly, we observed the 3′ UTR-length mediated repression of our reporters in three different human cell lines (A549, Flp-In T-REx 293 and HeLa cells), indicating the use of this pathway is widespread. However, the strength of the effect varied between the cell types. It will be interesting to determine how widely used this decay pathway is, both among different cell types and among different species.

Because these discoveries were made using randomly generated sequences, it was important to determine if a similar effect could be seen using endogenous sequences. Using concatenated sequences from endogenous human 3′ UTRs, we demonstrate that endogenously-derived 3′ UTR reporters are also affected by 3′ UTR length. However, the effect is weaker than was observed for random 3′ UTR mimics. In fact, the endogenously-derived 3′ UTR reporters show no length-effect in one of the cell lines (Flp-In T-Rex 293 cells). Two factors likely contribute to this result: First, the length-dependent decay pathway is much weaker in Flp-In T-Rex 293 cells than in the other two cell lines when measured using the random 3′ UTR mimics. Second, one of the endogenous 3′ UTR pieces clearly contains repressive sequences that are much stronger in Flp-In T-Rex 293 cells. Already destabilized transcripts may be less susceptible to additional repression, as was demonstrated by the reduced effect of 3′ UTR length on the destabilized yGFP reporters. Therefore, these repressive sequences may mask the effect of 3′ UTR length for a portion of the concatenated constructs.

Though the parameters discussed above may provide a partial explanation, the difference between our random 3′ UTR mimics and 3′ UTRs derived from endogenous sequences brings up the question of how 3′ UTR length is being recognized to trigger decay. Is the decay pathway genuinely being triggered by the distance between the stop codon and the polyA tail or is there some other feature that correlates with 3′ UTR length that

is being recognized? In other words, could this decay pathway be another surveillance mechanism, similar to NMD, that recognizes aberrant transcripts and that is triggered by our random 3′ UTR mimics? The fact that 3′ UTR regions previously determined to rescue transcripts from NMD are also able to, at least partially, rescue reporters from this novel pathway suggests that the two mechanisms may also share a role in mRNA surveillance. To begin to determine if our random 3′ UTRs mimics are activating a novel surveillance mechanism, a thorough analysis of the difference between the random 3′ UTR mimics and endogenous 3′ UTR sequences will be necessary.

The biggest advantage of using random 3′ UTR mimics is that it allowed us to uncouple 3′ UTR length from the effect of *cis*-regulatory sequence elements, which is a complicating factor in all analyses of the effect of 3′ UTR length. The use of artificial sequences to mimic or alter 3′ UTR length without the presence of regulatory motifs is not new. This has, for example, been done by inverting endogenous 3′ UTR sequences [9] or by the addition of unrelated eukaryotic or prokaryotic sequence to the 3′ UTR [148, 145]. However, in none of these examples was the absence of regulatory motifs in the artificial 3′ UTRs experimentally determined and in the latter two examples the sequences are likely to differ substantially in nucleotide composition from regular 3′ UTR sequence. The use of multiple randomly generated sequences, approximating the nucleotide composition of endogenous 3′ UTRs, allows us to overcome these shortcomings.

Despite the advantage the use of random sequences provides, our approach does have some shortcomings. For example, though we did control for nucleotide composition, we did not perfectly capture dinucleotide composition, which is also a distinctive feature of 3′ UTRs [201]. In addition, the presence of regulatory sequences that bind *trans*-factors may also be a defining feature of 3′ UTRs. Therefore, it is possible that our random 3′ UTR mimics are being recognized as aberrant due to either an unusual dinucleotide composition or a lack of bound *trans*-factors. If this is the case, future examination of our

random 3′ UTR mimics and how they differ from endogenous 3′ UTR sequences may reveal important information on how a 3′ UTR is recognized as normal or aberrant.

Even if this novel 3′ UTR-length dependent decay pathway is primarily a surveillance mechanism targeting 3′ UTRs based on an unknown non-typical parameter, it would still be likely to also regulate endogenous transcripts under certain conditions. This would parallel the NMD pathway, which is primarily a surveillance mechanism but has been coopted by cells to regulate non-defective mRNAs. Therefore, the information gleamed from further study of random 3′ UTR mimics may help us not only increase our understanding of RNA surveillance but also our understanding of 3′ UTR-mediated regulation of normal genes.

3.5 Materials and Methods

3′ UTR reporter constructs

Random 400mer 3′ UTR mimic luciferase reporter: The sequence of the random 400mer 3′ UTR mimics were generated to mimic the nucleotide composition of human 3′ UTRs (A:27%, U:29%, C:22%, G:22%) by making the probability of each base at each position mirror its occurrence rate in human 3′ UTRs. Any 400mer sequence containing target sites for the top 20 expressed miRNAs in the cell lines used, and any containing restriction sites used for cloning, were filtered out. Nine random 400mer sequences were chosen and synthesized as gBlocks (IDT), with a SpeI restriction site 5 and NheI and NotI restriction sites 3 of the sequence. Each 400mer was inserted downstream of the firefly luciferase coding sequence and upstream of the SV40 late poly(A) signal of Promega's pmirGLO Dual-Luciferase miRNA Target Expression vector, using the NheI and NotI vector restriction sites.

Concatenating to generate larger 3′ UTR mimics: During the cloning of the random 400mer 3′ UTR mimics the original NheI site is disrupted. However, the insert adds a new NheI site downstream of the random 400mer sequence and upstream of the NotI site. This allows for sequential cloning of random 400mers using the same restriction sites. Random 400mer 3′ UTR mimic plasmids are digested with NheI and NotI and another random 400mer insert is cloned in downstream of the first, again destroying the original NheI site and adding a new one downstream. See Table 3.1 for full list of concatamers.

Plasmids to test SpeI/NheI residual cloning motif: The residual cloning sequence between two random 400mers in two random 800mer concatamers (BE and BF) were mutated using the QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies) according to manufacturers instructions. Mutated 800mers were sequence verified, PCR extracted and cloned into a fresh pmirGLO vector. The residual cloning motif was added to two previously designed random 100mer 3′ UTR mimic sequences [125] and oligonucleotides designed to generate the sequences and clone them into pmirGLO using SpeI/NheI and NotI as described above.

Random 3′ UTR mimic GFP reporters: random 400mers and their concatamers were isolated from luciferase reporters using PCR, replacing the restriction sites with NotI and PmeI. These were then cloned between NotI and PmeI restriction sites in a GFP/dsRed pEF5/FRT/V5-D-TOPO derived plasmid [58].

No-insert control reporter: The no-3′ UTR control was generated by digesting the pmirGLO vector with NheI and NotI, or the GFP vector with NotI and PmeI, and using T4 DNA polymerase to create blunt ends, which were then ligated.

yGFP plasmid cloning: The GFP open reading frame in each random 3′ UTR mimic reporter was replaced with a yeast codon-optimized GFP, PCR amplified from a pFA6a-GFP-Trp1 plasmid generously supplied by the Emr laboratory. A start codon was added

to the pFA6a-GFP-Trp1-derived GFP during PCR amplification.

Endogenous human 3′ UTR luciferase reporters: Fragments from endogenous human 3′ UTRs had already been cloned from human genomic DNA and into reporter plasmids. These were PCR amplified from plasmids, adding SpeI upstream and NheI and NotI downstream, and cloned into pmirGLO as described for random 3′ UTR mimics. Concatenation also performed using method described above. See Table 3.2 for full list of 3′ UTRs used and their concatamers.

NMD escape 3′ UTR region reporters: The first 200nts of the human TRAM1 and VAMP3 3′ UTRs was isolated from genomic DNA using nested PCR, adding a PmeI site upstream and NheI and NotI sites downstream. These were cloned into pmirGLO using PmeI and NotI sites. The resulting plasmid was digested with NheI and NotI to add the SMG5 3′ UTR or 3′ UTR mimics downstream. The SMG5 3′ UTR was cloned from genomic DNA using nested PCR adding SpeI restriction site upstream and NheI and NotI sites downstream.

Plasmids for *in vitro* transcription: Random 3′ UTR mimics were cloned downstream of Firefly luciferase in a pUC18-derived *in vitro* transcription vector (described in [14]) using XhoI and BglII restriction sites (added to inserts with PCR).

In all cases inserts and cloning junctions were verified using Sanger sequencing.

Dual-luciferase reporter assays

Plasmid transfections: Cells were seeded in 10% FBS DMEM (Gibco), 24 hours before transfection at the following densities in 24-well plates: 75,000 cells/well for A549 cells, 150,000 cells/well for FlpIn293 cells and 50,000 cells/mL for HeLa cells. Plasmids (equimolar amounts, 14-20ng/well depending on plasmid size) were transfected into

cells, along with carrier DNA (100-140 ng/well pUC19), using Lipofectamine 2000 (Invitrogen) following the manufacturers protocols. Cells were harvested 30 h post-transfection by removing the media, washing once with 1 PBS, and freezing at −80°C. Luciferase assays were performed as before [125]. Firefly luciferase values were first normalized to Renilla values for each individual well to control for transfection efficiency, and then scaled relative to the normalized firefly values for the no-insert control (or to unaltered construct when testing mutations). The error bars were estimated as the nonparametric equivalent to one standard deviation (~68% of the data is within the error bars, i.e. the middle 7 datapoints for $n = 9$).

RNA transfection: Isolated *in vitro* transcribed RNAs were transfected as described above for plasmids with the following changes: 1) 10-50ng of Firefly RNA with random 3′ UTR mimics along with 5 ng of Renilla luciferase RNA. 400mer and 800mer pairs (B and BE, F and FE) were transfected in equimolar amounts. 2) Cells were harvested after 13 hours. 3) After normalization of Firefly to Renilla, the resulting values were normalized to the 400mer of the pair.

Making integrated GFP cell lines

Flp-In T-REx 293 cells (Life Technologies) were maintained in 10% FBS and 1% penicillin/streptomycin (P/S) supplemented DMEM. Cells were seeded at 300,000 cells/mL into 6-well plates using antibiotic free media (DMEM with 10% FBS) and transfected with 250 ng of GFP plasmid and 625 ng of pOG44 (encoding FLP recombinase) using Lipofectamine 2000 (Invitrogen). Media was replaced the following day with DMEM with 10% FBS and 1% P/S (complete DMEM). Selection for integrated cells started 48 hours post transfection by passaging cells into 10 cm plates into complete DMEM with 125 µg/mL hygromycin. Selection continued, with media changes every 3-4 days, until colonies were visible. Colonies were then counted (at least 10 colonies for each cell line) before being

passaged into T25 flasks, expanded and frozen.

Flow cytometry

GFP fluorescence of Flp-In T-REx 293 cells expressing integrated GFP reporters was quantified in 10,000 cells using a BD LSRII flow cytometer using DiVa software (BD Biosciences) and analyzed using FlowJo (TreeStar).

RNAi knockdown

siRNA transfections: HeLa cells were seeded at 75,000 cells/mL in 10cm culture dishes in antibiotic free DMEM with 10% FBS and grown for 24 hours. Cells were then transfected with 30 nM siRNA (UPF1: 5-GAUGCAGUUCCGCUCCAUUdTdT-3, 5-AAUGGAGCGGAACUGCAUCdTdT-3 [194]; I2.2: 5-UAAAAAUCGCGUGGAUUAAUG-3, 5-UUAAUUUACGCGGUUUUUAUU-3; I3.2 5-UAAUUUAACGCGGGUUUUAUC-3, 5-UAAAAUUCGCGUUGAGUUAAG-3) using the RNAiMAX transfection reagent (Invitrogen) according to manufacturers instructions. Transfected cells were seeded the following day on 24-well plates for plasmid transfection for dual luciferase assays (see above).

shRNA transductions were performed exactly as before [125], except using a higher MOI of 10.

qRT-PCR

RNA was isolated from tissue culture cells using the Trizol reagent (Thermo Fisher Scientific). cDNA was generated using a random nonamer (dN9) primer and the RevertAid

Reverse transcriptase (Thermo Scientific) at 42°C for 60 min; the reaction was heat inactivated at 70°C for 10 min. qPCR reactions were performed using Taq polymerase and SYBR Green (Life Technologies) as the detection agent. GAPDH was used as a housekeeping gene to which to normalize. Each qPCR reaction was done in triplicate.

Half-life measurements

Flp-In T-REx 293 cells with integrated GFP reporters were seeded at 350,000 cells/mL into 6 well plates in complete DMEM and let grow for ~24 hours. To stop transcription, media was replaced with complete media containing 2.5 µg/mL Actinomycin D (Life Technologies). Cells were harvested before, and at 30 minutes, 1 hour, 2 hours, 4 hours and 6 hours after, the addition of actinomycin media by removing the media, adding Trizol reagent and freezing at -80°C until ready to isolate RNA.

***In vitro* transcription**

In vitro transcription plasmids were linearized with either BsaI or HindIII to generate templates for mRNAs containing either a regular polyA tail or a protected polyA tail. To make a template for a control Renilla mRNA, a previously described Renilla luciferase *in vitro* transcription plasmid [14] was linearized with BamHI. *In vitro* transcription was performed with T3 RNA polymerase (Agilent Technologies) at 37°C for 3 hours and then DNase treated with DNase I (Roche) for 1 hour. Capping reactions were performed during *in vitro* transcription by including m7GpppG/GTP (Jena Bioscience) a 2:1 molar ratio. RNA was isolated from the reactions using phenol/chloroform/isoamylalcohol extraction.

CHAPTER 4

CHAPTER 4

4.1 Conclusions

The overarching goal of this thesis was to answer fundamental questions about the workings of 3′ UTRs, to further our understanding of their role in post-transcriptional gene regulation. The three major questions I posed were: 1) How densely packed are 3′ UTRs with regulatory information? 2) How commonly do regulatory *cis*-elements within 3′ UTRs interact with one another? 3) What is the role of 3′ UTR length in determining gene expression levels? Though many questions still remain, and the questions I posed still merit further exploration, the answers provided in this thesis contribute to our understanding of 3′ UTR-mediated post-transcriptional gene regulation.

To answer the first two questions, about the density of regulatory elements and the ubiquity of interactions between them, I performed a systematic study of two conserved 3′ UTRs, those of *Hmga2* and *PIM1*. By dividing the 3′ UTRs into small segments and measuring the regulatory effect of each of the segments in reporter assays, I was able to determine the approximate proportion of the 3′ UTR that contains functional regulatory information. I also generated larger segments and used a comparison between those and the constituent smaller segments to identify cases of interactions between elements within adjacent segments.

For both *Hmga2* and *PIM1*, about 60% of the segments significantly affected reporter expression (35 of 58 50mer segments of *Hmga2* and 8 of 13 100mer segments of *PIM1*), giving an idea of the density of regulatory information within 3′ UTRs. Some additional modeling is necessary to estimate the percentage of 3′ UTR sequence that is directly involved in regulating gene expression; the vast majority of regulatory elements are much

smaller than 50 nts, meaning only a portion of each segment is likely to contain regulatory information and there is space for more than one *cis*-element *per se*gment. Using a simple binomial distribution model to estimate the likely total number of regulatory *cis*-elements, given the percentage of *Hmga2* 3′ UTR 50mers that contain at least one *cis*-element, we get about 49 *cis*-elements, making up around 16.4% of the *Hmga2* 3′ UTR (assuming 10nts per element). This matches quite well the estimate arrived at by global PhastCons conservation analysis of 3′ UTRs [6].

Another striking result from my analysis of 3′ UTR segments was the large number of segments that had a positive effect on reporter expression. This is in opposition to the prevailing view of 3′ UTRs as having mainly a repressive effect. This result was not limited to a single condition, but was consistent across two separate 3′ UTRs. In addition, the number of positive regions in the *Hmga2* 3′ UTR was quite consistent in four different cell lines and in 3′ UTRs from three different vertebrate species. Other studies have also started to see similar results [58, 143], indicating that the large number of positive elements may be true of many 3′ UTRs.

When I expanded the analysis of the two 3′ UTRs to look for interactions between regulatory elements within adjacent 3′ UTR segments, I found very little evidence of any interactions. For *Hmga2*, which I studied in more detail, this was true going from 50mer to 100mer segments, from 100mer to 200mer segments and from 200 to 400mer segments. Even using 50mer segments to model the effect of 400mer segments produced incredibly accurate results. These data strongly indicate that there are very few interactions between regulatory elements within this size range. However, they do not exclude longer-range interactions, such as those observed by Sugimoto et al. when exploring the double-stranded RNA binding sites of Staufen [55]. In fact, I did find evidence that such an interaction may take place. A regulatory element identified in the body of the 3′ UTR switched its function from repressive in the context of the full length *Hmga2* 3′ UTR to activating when the

first and last 400 nts of the 3′ UTR were not included.

Both the *Hmga2* and *PIM1* 3′ UTRs were very repressive, when the full length 3′ UTR was compared to a no-3′ UTR control reporter, in contradiction to the combined regulatory effect measured in the small segments. This could be explained by assuming a large number of positive sites switch function, as at least one element did in the *Hmga2* 3′ UTR. However, this seems quite unlikely to happen on a large enough scale. Another potential factor is a possible repressive effect of 3′ UTR length, which is not controlled for when comparing these two relatively long 3′ UTRs with a no-3′ UTR control. This, in addition to the observation that the random-sequence length-matched controls we used to normalize our 3′ UTR segments showed the beginning of a negative trend in expression in response to 3′ UTR length, convinced me to take a closer look at the role 3′ UTR size plays in post-transcriptional gene regulation.

To study the effect of 3′ UTR length on gene expression, it is necessary to separate the effect of regulatory sequences from the *per se* effect of 3′ UTR length. I accomplished this by using multiple randomly generated sequences as 3′ UTR mimics, approximating the nucleotide composition of human 3′ UTRs. Using these mimics, I discovered a very strong negative effect of 3′ UTR length on reporter expression. The repression of long 3′ UTR reporters occurred through an NMD-independent decay mechanism which affects the half-life of mature mRNA transcripts in a deadenylation-independent manner. This is the first identification, to the best of my knowledge, of an NMD-independent pathway through which 3′ UTR length regulates gene expression. While I also observed an effect of 3′ UTR length on reporters with endogenously-derived 3′ UTRs, the effect was weaker.

4.2 Future Directions

Like all scientific pursuits, these studies leave many more questions to explore. For example, while I identified regions that contain regulatory information, I was only able to narrow down the exact sequence responsible for a small number of those regions. In addition, it remains to be seen how generally applicable my *cis*-element density and interaction estimates are to a wider assortment of 3′ UTRs. As the approach taken in chapter 2 is very labor intensive and expensive to perform on a large scale, expanding the scope of this research to either increase the resolution or study many more 3′ UTRs will require the implementation of a high-throughput screening technique. One such technique is currently in use the Grimson laboratory [58] that may easily be adapted for this purpose. This screening technique involves GFP reporters with thousands of different 3′ UTR segments cloned in bulk into the 3′ UTR. These are integrated into a single genomic locus in Flp-In T-REx 293 cells, along with a dsRed control reporter, and fluorescence activated cell sorting (FACS) used to isolate cells expressing GFP at varying levels (and dsRed at constant levels). High-throughput sequencing is then used to identify the 3′ UTR segments at each GFP level and this information used to calculate the regulatory potential of those segments.

Another avenue of further research involving the *Hmga2* 3′ UTR is the identification of *trans*-factors that bind to identified regulatory *cis*-elements. One possible approach is to use RNA-IPs followed by mass spectrometry, comparing intact 3′ UTR segments with mutated versions for the sites of interest. Another approach would be an RNA interference (RNAi) screen to identify proteins whose knockdown affects *Hmga2* reporter expression. In this approach, stable cell lines expressing the wild-type and mutant versions of *Hmga2* attached to different fluorescent reporters are subjected to lentiviral short hairpin RNA (shRNA) infection to knock down a cadre of RNA binding proteins. Following infection cells with altered wild-type over mutant reporter expression ratios are isolated

using FACS and the shRNAs expressed by those cells identified with high-throughput sequencing. Candidate *trans*-factors that regulate expression through the *cis*-elements in question may then be identified as RBPs for whom two or more shRNAs consistently result in altered reporter ratios.

Finally, the functional switch I discovered for a *cis*-element in the *Hmga2* 3′ UTR warrants further study. Specifically, it would be important to know if there is a physical interaction between the *cis*-element and the interacting regions at the 3′ UTR ends or if those edge regions contain elements that change the cellular environment. The edge regions could change the cellular environment by changing the subcellular localization of the mRNA or its sequestration into granules. The first step to elucidating this would be to determine if the localization of *Hmga2* transcripts is changed by removing the edge regions. To do this one could add aptamers that bind multiple copies of GFP (e.g. MS2 aptamers and MS2-GFP chimeric proteins) to an mRNA containing the *Hmga2* 3′ UTR with and without those edge sequences. One could then use microscopy to monitor the transcripts localization or sequestration into granules. If localization is not affected, the specific RNA regions and *trans*-factors involved in a physical interaction need to be identified using mutational analysis and the RNA-IP strategy described above.

For the role of 3′ UTR length in regulating gene expression, many questions remain unanswered. Specifically pertaining to the decay mechanism identified in chapter 3, the nucleases that degrade the long 3′ UTR transcripts remain to be identified. In addition, the mechanism by which the length of the 3′ UTR is measured and how decay enzymes are recruited has not been elucidated. Finally, the biological impact of this pathway needs to be determined. An experiment is already underway to determine if long 3′ UTR transcripts are degraded by 5-to-3 or 3-to-5 exonucleolytic decay. We are knocking down critical components of the Xrn1 (5-to-3) or exosome (3-to-5) complexes and measuring the effect on integrated yGFP reporters with long or short random 3′ UTR mimics. As

we have excluded deadenylation as a mechanism, the involvement of 3-to-5 decay by the exosome would indicate the possible involvement of endonucleases. By contrast, the involvement of Xrn1 only would implicate decapping as the major trigger of decay.

The mechanism by which the length of the 3′ UTR is measured in more difficult to determine. There are two major avenues of research I would consider: 1) As the endogenously-derived 3′ UTRs were less strongly affected by 3′ UTR length, identifying differences between our random 3′ UTR mimics and endogenous 3′ UTR sequences may provide useful hypotheses. 2) Using an RNAi screen as described above to identify proteins whose knockdown reduces the difference in expression between reporters with long and short 3′ UTRs. Either approach would provide important insights into how long 3′ UTR are targeted, and the second approach might provide further details about the overall decay mechanism.

Once the identity of crucial *trans*-factors has been established, the biological impact of this decay pathway can be determined. The conservation of this pathway may then be inferred from the conservation of the specific proteins necessary to trigger decay through this pathway. The next step would be to identify likely endogenous target by monitoring changes to the transcriptome after overexpression or knockdown of critical components of the decay pathway. Finally, the physiological consequences of disrupting this pathway can be determined by knock-out experiments in model organisms, such as mice, depending on the level of conservation.

Another avenue of inquiry would be the apparent mutual exclusivity of NMD and the novel pathway. What is it about our 3′ UTR mimics that make them targets of the novel pathway rather than NMD? Again, an answer may be found by exploring the difference between our random 3′ UTR mimics and known targets of NMD that are targeted due to 3′ UTR length. UPF1 is known to prefer to bind GC-rich sequences [193], so a first question might be if targets of NMD are more GC-rich than targets of the novel pathway.

The fact that shorter 3′ UTR mimics did seem to be affected by NMD disruption indicates that our 3′ UTR mimics can be targeted by NMD. My hypothesis is that as the 3′ UTR mimics get longer, decay by the novel pathway outcompetes NMD. Once the identity of crucial components of the novel pathway are identified, it will be possible to test this hypothesis by inhibiting the novel pathway and determining if the long 3′ UTR mimics then become targets of NMD.

Finally, the identification of a second decay mechanism, after NMD, that is triggered by 3′ UTR length indicates that other regulatory pathways may also be affected by 3′ UTR length. Though I found no evidence of such pathways affecting the random 3′ UTR mimics in the particular systems we used, it is possible that other pathways may also be outcompeted by the identified decay pathway, as seems to be the case for NMD. Because of this, the identification of such mechanisms will likely be serendipitous, as the appropriate conditions and targets are unknown. Perhaps, new pathways may be identified by creating 3′ UTR mimics whose parameters, such as nucleotide or dinucleotide composition or secondary structures, are varied.

Overall, 3′ UTRs remain an exciting subject of research with many fundamental properties still up for debate. For example, while my findings in chapter 2 suggest *cis*-element interactions are rare, global computational studies seem to indicate that they may be widespread, at least for certain *cis*-elements [127, 126, 129, 128, 102]. Further study is needed to reconcile these different findings. In addition, we are only just beginning to understand the role of 3′ UTR length in gene regulation. Beyond the specific topics of research focused on in this thesis, the role and extent of secondary or tertiary structures within 3′ UTRs is still widely debated and the role of RNA modifications is just starting to be elucidated.

BIBLIOGRAPHY

[1] Li, J. J., Bickel, P. J., and Biggin, M. D. (2014) System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ*, **2**, e270.

[2] Elkon, R., Zlotorynski, E., Zeller, K. I., and Agami, R. (2010) Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics*, **11**, 259.

[3] Prez-Ortn, J. E., Alepuz, P. M., and Moreno, J. (2007) Genomics and gene transcription kinetics in yeast. *Trends in genetics: TIG*, **23**(5), 250–257.

[4] Moor, C. H. d., Meijer, H., and Lissenden, S. (2005) Mechanisms of translational control by the 3 UTR in development and differentiation. *Seminars in Cell & Developmental Biology*, **16**(1), 49–58.

[5] Zhao, W., Blagev, D., Pollack, J. L., and Erle, D. J. (2011) Toward a systematic understanding of mRNA 3' untranslated regions. *Proceedings of the American Thoracic Society*, **8**(2), 163–166.

[6] Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., and Haussler, D. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, **15**(8), 1034–1050.

[7] Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, **20**(1), 110–121.

[8] Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008) Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science*, **320**(5883), 1643–1647.

[9] Mayr, C. and Bartel, D. P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**(4), 673–684.

[10] Mitchell, S. F. and Parker, R. (2014) Principles and Properties of Eukaryotic mRNPs. *Molecular Cell*, **54**(4), 547–558.

[11] Szostak, E. and Gebauer, F. (2013) Translational control by 3-UTR-binding proteins. *Briefings in Functional Genomics*, **12**(1), 58–65.

[12] Sampath, P., Mazumder, B., Seshadri, V., and Fox, P. L. (2003) Transcript-selective translational silencing by gamma interferon is directed by a novel structural element in the ceruloplasmin mRNA 3′ untranslated region. *Molecular and Cellular Biology,* **23**(5), 1509–1519.

[13] Kapasi, P., Chaudhuri, S., Vyas, K., Baus, D., Komar, A. A., Fox, P. L., Merrick, W. C., and Mazumder, B. (2007) L13a blocks 48S assembly: role of a general initiation factor in mRNA-specific translational control. *Molecular Cell,* **25**(1), 113–126.

[14] Geissler, R., Simkin, A., Floss, D., Patel, R., Fogarty, E. A., Scheller, J., and Grimson, A. (2016) A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes & Development,* **30**(9), 1070–1085.

[15] Eichhorn, S., Guo, H., McGeary, S., Rodriguez-Mias, R., Shin, C., Baek, D., Hsu, S.-h., Ghoshal, K., Villn, J., and Bartel, D. (2014) mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. *Molecular Cell,* **56**(1), 104–115.

[16] Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research,* **42**(D1), D68–D73.

[17] Garcia, D. M., Baek, D., Shin, C., Bell, G. W., Grimson, A., and Bartel, D. P. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nature Structural & Molecular Biology,* **18**(10), 1139–1146.

[18] Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research,* **19**(1), 92–105.

[19] Holt, C. E. and Bullock, S. L. (2009) Subcellular mRNA localization in animal cells and why it matters. *Science (New York, N.Y.),* **326**(5957), 1212–1216.

[20] Lcuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., and Krause, H. M. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell,* **131**(1), 174–187.

[21] Cajigas, I. J., Tushev, G., Will, T. J., tom Dieck, S., Fuerst, N., and Schuman, E. M. (2012) The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron,* **74**(3), 453–466.

[22] Gumy, L. F., Yeo, G. S. H., Tung, Y.-C. L., Zivraj, K. H., Willis, D., Coppola, G., Lam,

B. Y. H., Twiss, J. L., Holt, C. E., and Fawcett, J. W. (2011) Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. *RNA,* **17**(1), 85–98.

[23] De Domenico, E., Owens, N. D. L., Grant, I. M., Gomes-Faria, R., and Gilchrist, M. J. (2015) Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing. *Developmental Biology,* **408**(2), 252–268.

[24] Mili, S., Moissoglu, K., and Macara, I. G. (2008) Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature,* **453**(7191), 115–119.

[25] Condeelis, J. and Singer, R. H. (2005) How and why does -actin mRNA target?. *Biology of the Cell,* **97**(1), 97–110.

[26] Besse, F. and Ephrussi, A. (2008) Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nature Reviews. Molecular Cell Biology,* **9**(12), 971–980.

[27] Shahbabian, K. and Chartrand, P. (2012) Control of cytoplasmic mRNA localization. *Cellular and molecular life sciences: CMLS,* **69**(4), 535–552.

[28] Martin, K. C. and Ephrussi, A. (2009) mRNA localization: gene expression in the spatial dimension. *Cell,* **136**(4), 719–730.

[29] Berkovits, B. D. and Mayr, C. (2015) Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature,* **522**(7556), 363–367.

[30] Tsvetanova, N. G., Klass, D. M., Salzman, J., and Brown, P. O. (2010) Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PloS One,* **5**(9).

[31] Baltz, A. G., Munschauer, M., Schwanhusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell,* **46**(5), 674–690.

[32] Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell,* **149**(6), 1393–1406.

[33] Kwon, S. C., Yi, H., Eichelbaum, K., Fhr, S., Fischer, B., You, K. T., Castello, A., Krijgsveld, J., Hentze, M. W., and Kim, V. N. (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nature Structural & Molecular Biology*, **20**(9), 1122–1130.

[34] Mitchell, S. F., Jain, S., She, M., and Parker, R. (2013) Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*, **20**(1), 127–133.

[35] Liepelt, A., Naarmann-de Vries, I. S., Simons, N., Eichelbaum, K., Foehr, S., Archer, S. K., Castello, A., Usadel, B., Krijgsveld, J., Preiss, T., Marx, G., Hentze, M. W., Ostareck, D. H., and Ostareck-Lederer, A. (2016) Identification of RNA-binding proteins in macrophages by interactome capture. *Molecular & cellular proteomics: MCP,*.

[36] Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., Krijgsveld, J., and Hentze, M. W. (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications*, **6**, 10127.

[37] Ciela, J. (2006) Metabolic enzymes that bind RNA: yet another level of cellular regulatory network?. *Acta Biochimica Polonica*, **53**(1), 11–32.

[38] Castello, A., Hentze, M. W., and Preiss, T. (2015) Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends in endocrinology and metabolism: TEM*, **26**(12), 746–757.

[39] Dabo, S. and Meurs, E. F. (2012) dsRNA-dependent protein kinase PKR and its role in stress, signaling and HCV infection. *Viruses*, **4**(11), 2598–2635.

[40] Wek, S. A., Zhu, S., and Wek, R. C. (1995) The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. *Molecular and Cellular Biology*, **15**(8), 4497–4506.

[41] Habjan, M. and Pichlmair, A. (2015) Cytoplasmic sensing of viral nucleic acids. *Current Opinion in Virology*, **11**, 31–37.

[42] Mazurek, S., Hugo, F., Failing, K., and Eigenbrodt, E. (1996) Studies on associations of glycolytic and glutaminolytic enzymes in MCF-7 cells: role of P36. *Journal of Cellular Physiology*, **167**(2), 238–250.

[43] Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E. S., and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3 UTRs by comparison of several mammals. *Nature*, **434**(7031), 338–345.

[44] Yoon, K., Ko, D., Doderer, M., Livi, C. B., and Penalva, L. O. F. (2008) Over-represented sequences located on 3′ UTRs are potentially involved in regulatory functions. *RNA biology,* **5**(4), 255–262.

[45] Cor, D., Di Cunto, F., Caselle, M., and Provero, P. (2007) Identification of candidate regulatory sequences in mammalian 3′ UTRs by statistical analysis of oligonucleotide distributions. *BMC bioinformatics,* **8**, 174.

[46] Dassi, E., Zuccotti, P., Leo, S., Provenzani, A., Assfalg, M., DOnofrio, M., Riva, P., and Quattrone, A. (2013) Hyper conserved elements in vertebrate mRNA 3-UTRs reveal a translational network of RNA-binding proteins controlled by HuR. *Nucleic Acids Research,* **41**(5), 3201–3216.

[47] Cohen, J. E., Lee, P. R., and Fields, R. D. (2014) Systematic identification of 3′-UTR regulatory elements in activity-dependent mRNA stability in hippocampal neurons. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* **369**(1652).

[48] Goodarzi, H., Najafabadi, H. S., Oikonomou, P., Greco, T. M., Fish, L., Salavati, R., Cristea, I. M., and Tavazoie, S. (2012) Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature,* **485**(7397), 264–268.

[49] Ray, D., Kazan, H., Chan, E. T., Castillo, L. P., Chaudhry, S., Talukder, S., Blencowe, B. J., Morris, Q., and Hughes, T. R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotechnology,* **27**(7), 667–670.

[50] Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., and Burge, C. B. (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Molecular Cell,* **54**(5), 887–900.

[51] Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell,* **141**(1), 129–141.

[52] Yoon, J.-H., De, S., Srikantan, S., Abdelmohsen, K., Grammatikakis, I., Kim, J., Kim, K. M., Noh, J. H., White, E. J. F., Martindale, J. L., Yang, X., Kang, M.-J., 3rd, W. H. W., Hooten, N. N., Evans, M. K., Becker, K. G., Tripathi, V., Prasanth, K. V., Wilson, G. M., Tuschl, T., Ingolia, N. T., Hafner, M., and Gorospe, M. (2014) PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nature Communications,* **5**.

[53] Mukherjee, N., Jacobs, N. C., Hafner, M., Kennington, E. A., Nusbaum, J. D., Tuschl, T., Blackshear, P. J., and Ohler, U. (2014) Global target mRNA specification and regulation by the RNA-binding protein ZFP36. *Genome Biology,* **15**(1), R12.

[54] Kenny, P. J., Zhou, H., Kim, M., Skariah, G., Khetani, R. S., Drnevich, J., Arcila, M. L., Kosik, K. S., and Ceman, S. (2014) MOV10 and FMRP regulate AGO2 association with microRNA recognition elements. *Cell reports,* **9**(5), 1729–1741.

[55] Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N. M., and Ule, J. (2015) hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature,* **519**(7544), 491–494.

[56] Oikonomou, P., Goodarzi, H., and Tavazoie, S. (2014) Systematic Identification of Regulatory Elements in Conserved 3 UTRs of Human Transcripts. *Cell Reports,* **7**(1), 281–292.

[57] Zhao, W., Pollack, J. L., Blagev, D. P., Zaitlen, N., McManus, M. T., and Erle, D. J. (2014) Massively parallel functional annotation of 3 untranslated regions. *Nature Biotechnology,* **32**(4), 387–391.

[58] Wissink, E. M., Fogarty, E. A., and Grimson, A. (2016) High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics,* **17**, 177.

[59] Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell,* **27**(1), 91–105.

[60] Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell,* **153**(3), 654–665.

[61] Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife,* **4**, e05005.

[62] Miller, M. A. and Olivas, W. M. (2011) Roles of Puf proteins in mRNA degradation and translation. *Wiley interdisciplinary reviews: RNA,* **2**(4), 471–492.

[63] Barreau, C., Paillard, L., and Osborne, H. B. (2005) AU-rich elements and associated factors: are there unifying principles?. *Nucleic Acids Research,* **33**(22), 7138–7150.

[64] Vlasova-St. Louis, I. and Bohjanen, P. R. (2014) Post-Transcriptional Regulation of Cytokine Signaling by AU-Rich and GU-Rich Elements. *Journal of Interferon & Cytokine Research,* **34**(4), 233–241.

[65] Masliah, G., Barraud, P., and Allain, F. H.-T. (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cellular and Molecular Life Sciences,* **70**(11), 1875–1895.

[66] Stefl, R., Oberstrass, F. C., Hood, J. L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R. B., and Allain, F. H.-T. (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell,* **143**(2), 225–237.

[67] Li, X., Quon, G., Lipshitz, H. D., and Morris, Q. (2010) Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA,* **16**(6), 1096–1107.

[68] Kim, H. H., Lee, S. J., Gardiner, A. S., Perrone-Bizzozero, N. I., and Yoo, S. (2015) Different motif requirements for the localization zipcode element of -actin mRNA binding by HuD and ZBP1. *Nucleic Acids Research,* **43**(15), 7432–7446.

[69] Heraud-Farlow, J. E. and Kiebler, M. A. (2014) The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. *Trends in Neurosciences,* **37**(9), 470–479.

[70] Park, E. and Maquat, L. E. (2013) Staufen-mediated mRNA decay. *Wiley interdisciplinary reviews: RNA,* **4**(4), 423–435.

[71] Hake, L. E. and Richter, J. D. (1994) CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. *Cell,* **79**(4), 617–627.

[72] Huang, Y.-S., Kan, M.-C., Lin, C.-L., and Richter, J. D. (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *The EMBO Journal,* **25**(20), 4865–4876.

[73] Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D., and Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature,* **499**(7457), 172–177.

[74] Chen, L. (2013) Characterization and comparison of human nuclear and cytosolic editomes. *Proceedings of the National Academy of Sciences of the United States of America,* **110**(29), E2741–2747.

[75] Nishikura, K. (2010) Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual review of biochemistry,* **79**, 321–349.

[76] Gu, T., Buaas, F. W., Simons, A. K., Ackert-Bicknell, C. L., Braun, R. E., and Hibbs, M. A. (2012) Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. *PloS One,* **7**(3), e33720.

[77] Zhang, L., Yang, C.-S., Varelas, X., and Monti, S. (2016) Altered RNA editing in 3' UTR perturbs microRNA-mediated regulation of oncogenes and tumor-suppressors. *Scientific Reports,* **6**, 23226.

[78] Wang, I. X., So, E., Devlin, J. L., Zhao, Y., Wu, M., and Cheung, V. G. (2013) ADAR regulates RNA editing, transcript stability, and gene expression. *Cell Reports,* **5**(3), 849–860.

[79] Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015) RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science,* **349**(6252), 1115–1120.

[80] Smith, H. C., Bennett, R. P., Kizilyer, A., McDougall, W. M., and Prohaska, K. M. (2012) Functions and regulation of the APOBEC family of proteins. *Seminars in Cell & Developmental Biology,* **23**(3), 258–268.

[81] Koito, A. and Ikeda, T. (2012) Apolipoprotein B mRNA-editing, catalytic polypeptide cytidine deaminases and retroviral restriction. *Wiley interdisciplinary reviews: RNA,* **3**(4), 529–541.

[82] Prohaska, K. M., Bennett, R. P., Salter, J. D., and Smith, H. C. (2014) The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley interdisciplinary reviews: RNA,* **5**(4), 493–508.

[83] Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S., and Papavasiliou, F. N. (2011) Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nature Structural & Molecular Biology,* **18**(2), 230–236.

[84] Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., Len-Ricardo, B. X., Engreitz, J. M., Guttman, M., Satija, R., Lander, E. S., Fink, G., and Regev, A. (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell,* **159**(1), 148–162.

[85] Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert,

W. V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature,* **515**(7525), 143–146.

[86] Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., Suter, C. M., and Preiss, T. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research,* **40**(11), 5023–5033.

[87] Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013) Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biology,* **14**(11), 215.

[88] Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., and Rechavi, G. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature,* **485**(7397), 201–206.

[89] Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012) Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3 UTRs and near Stop Codons. *Cell,* **149**(7), 1635–1646.

[90] Licht, K. and Jantsch, M. F. (2016) Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *The Journal of Cell Biology,* **213**(1), 15–22.

[91] Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014) Gene expression regulation mediated through reversible mA RNA methylation. *Nature Reviews. Genetics,* **15**(5), 293–306.

[92] Maity, A. and Das, B. (2016) N6-methyladenosine modification in mRNA: machinery, function and implications for health and diseases. *FEBS Journal,* **283**(9), 1607–1630.

[93] Ke, S., Alemu, E. A., Mertens, C., Gantman, E. C., Fak, J. J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M. J., Park, C. Y., Vgb, C. B., Kusnierczyk, A., Klungland, A., Darnell, J. E., and Darnell, R. B. (2015) A majority of m6A residues are in the last exons, allowing the potential for 3 UTR regulation. *Genes & Development,* **29**(19), 2037–2053.

[94] Chen, T., Hao, Y.-J., Zhang, Y., Li, M.-M., Wang, M., Han, W., Wu, Y., Lv, Y., Hao, J., Wang, L., Li, A., Yang, Y., Jin, K.-X., Zhao, X., Li, Y., Ping, X.-L., Lai, W.-Y., Wu, L.-G., Jiang, G., Wang, H.-L., Sang, L., Wang, X.-J., Yang, Y.-G., and Zhou, Q. (2015) m6A RNA Methylation Is Regulated by MicroRNAs and Promotes Reprogramming to Pluripotency. *Cell Stem Cell,* **16**(3), 289–301.

[95] Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., Weng, X., Chen,

K., Shi, H., and He, C. (2015) N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell,* **161**(6), 1388–1399.

[96] Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., Ren, B., Pan, T., and He, C. (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature,* **505**(7481), 117–120.

[97] Alarcn, C. R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., and Tavazoie, S. F. (2015) HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell,* **162**(6), 1299–1308.

[98] Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nature Genetics,* **39**(10), 1278–1284.

[99] Long, D., Lee, R., Williams, P., Chan, C. Y., Ambros, V., and Ding, Y. (2007) Potent effect of target structure on microRNA function. *Nature Structural & Molecular Biology,* **14**(4), 287–294.

[100] Robins, H., Li, Y., and Padgett, R. W. (2005) Incorporating structure to predict microRNA targets. *Proceedings of the National Academy of Sciences of the United States of America,* **102**(11), 4006–4009.

[101] Ouyang, Z., Snyder, M. P., and Chang, H. Y. (2013) SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Research,* **23**(2), 377–387.

[102] HafezQorani, S., Lafzi, A., de Bruin, R. G., van Zonneveld, A. J., van der Veer, E. P., Son, Y. A., and Kazan, H. (2016) Modeling the combined effect of RNA-binding proteins and microRNAs in post-transcriptional regulation. *Nucleic Acids Research,* **44**(9), e83.

[103] Kierzek, E. and Kierzek, R. (2003) The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Research,* **31**(15), 4472–4480.

[104] Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature,* **518**(7540), 560–564.

[105] Strom, P., Heale, B. S. E., Snve, O., Aagaard, L., Alluin, J., and Rossi, J. J. (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Research,* **35**(7), 2333–2342.

[106] Broderick, J. A., Salomon, W. E., Ryder, S. P., Aronin, N., and Zamore, P. D. (2011) Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *RNA,* **17**(10), 1858–1869.

[107] Connerty, P., Ahadi, A., and Hutvagner, G. (2015) RNA Binding Proteins in the miRNA Pathway. *International Journal of Molecular Sciences,* **17**(1), 31.

[108] Iadevaia, V. and Gerber, A. P. (2015) Combinatorial Control of mRNA Fates by RNA-Binding Proteins and Non-Coding RNAs. *Biomolecules,* **5**(4), 2207–2222.

[109] Kouwenhove, M. v., Kedde, M., and Agami, R. (2011) MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Reviews Cancer,* **11**(9), 644–656.

[110] Cho, S. J., Zhang, J., and Chen, X. (2010) RNPC1 modulates the RNA-binding activity of, and cooperates with, HuR to regulate p21 mRNA stability. *Nucleic Acids Research,* **38**(7), 2256–2267.

[111] Kedde, M., Kouwenhove, M. v., Zwart, W., Vrielink, J. A. F. O., Elkon, R., and Agami, R. (2010) A Pumilio-induced RNA structure switch in p27-3 UTR controls miR-221 and miR-222 accessibility. *Nature Cell Biology,* **12**(10), 1014–1020.

[112] Miles, W. O., Tschp, K., Herr, A., Ji, J.-Y., and Dyson, N. J. (2012) Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes & Development,* **26**(4), 356–368.

[113] Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y., Jiang, L., Cai, Z., Sun, H., Zhang, K., Zhang, Y., Chen, J., and Fu, X.-D. (2013) Direct Conversion of Fibroblasts to Neurons by Reprogramming PTB-Regulated microRNA Circuits. *Cell,* **152**(1-2), 82–96.

[114] Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., Martindale, J. L., and Gorospe, M. (2009) HuR recruits let-7/RISC to repress c-Myc expression. *Genes & Development,* **23**(15), 1743–1748.

[115] Topisirovic, I., Siddiqui, N., Orolicki, S., Skrabanek, L. A., Tremblay, M., Hoang, T., and Borden, K. L. B. (2009) Stability of eukaryotic translation initiation factor 4E mRNA is regulated by HuR, and this activity is dysregulated in cancer. *Molecular and Cellular Biology,* **29**(5), 1152–1162.

[116] Tiedje, C., Ronkina, N., Tehrani, M., Dhamija, S., Laass, K., Holtmann, H., Kotlyarov, A., and Gaestel, M. (2012) The p38/MK2-driven exchange between tristetraprolin and HuR regulates AU-rich element-dependent translation. *PLoS genetics,* **8**(9), e1002977.

[117] Liu, L., Ouyang, M., Rao, J. N., Zou, T., Xiao, L., Chung, H. K., Wu, J., Donahue, J. M., Gorospe, M., and Wang, J.-Y. (2015) Competition between RNA-binding proteins CELF1 and HuR modulates MYC translation and intestinal epithelium renewal. *Molecular Biology of the Cell*, **26**(10), 1797–1810.

[118] Kedde, M., Strasser, M. J., Boldajipour, B., Vrielink, J. A. F. O., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., rom, U. A., Lund, A. H., Perrakis, A., Raz, E., and Agami, R. (2007) RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA. *Cell*, **131**(7), 1273–1286.

[119] Jafarifar, F., Yao, P., Eswarappa, S. M., and Fox, P. L. (2011) Repression of VEGFA by CA-rich element-binding microRNAs is modulated by hnRNP L. *The EMBO Journal*, **30**(7), 1324–1334.

[120] Zuccotti, P., Colombrita, C., Moncini, S., Barbieri, A., Lunghi, M., Gelfi, C., De Palma, S., Nicolin, A., Ratti, A., Venturin, M., and Riva, P. (2014) hnRNPA2/B1 and nELAV proteins bind to a specific U-rich element in CDK5R1 3′-UTR and oppositely regulate its expression. *Biochimica Et Biophysica Acta*, **1839**(6), 506–516.

[121] Young, L. E., Moore, A. E., Sokol, L., Meisner-Kober, N., and Dixon, D. A. (2012) The mRNA stability factor HuR inhibits microRNA-16 targeting of COX-2. *Molecular cancer research: MCR*, **10**(1), 167–180.

[122] Epis, M. R., Barker, A., Giles, K. M., Beveridge, D. J., and Leedman, P. J. (2011) The RNA-binding protein HuR opposes the repression of ERBB-2 gene expression by microRNA miR-331-3p in prostate cancer cells. *The Journal of Biological Chemistry*, **286**(48), 41442–41454.

[123] Kundu, P., Fabian, M. R., Sonenberg, N., Bhattacharyya, S. N., and Filipowicz, W. (2012) HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Research*, **40**(11), 5088–5100.

[124] Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., and Filipowicz, W. (2006) Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell*, **125**(6), 1111–1124.

[125] Kristjnsdttir, K., Fogarty, E. A., and Grimson, A. (2015) Systematic analysis of the Hmga2 3 UTR identifies many independent regulatory sequences and a novel interaction between distal sites. *RNA,*.

[126] Leibovich, L., Mandel-Gutfreund, Y., and Yakhini, Z. (2010) A structural-based statistical approach suggests a cooperative activity of PUM1 and miR-410 in human 3′-untranslated regions. *Silence*, **1**, 17.

[127] Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M., and Gerber, A. P. (2008) Comparative Analysis of mRNA Targets for Human PUF-Family Proteins Suggests Extensive Interaction with the miRNA Regulatory System. *PLoS ONE,* **3**(9), e3164.

[128] Jiang, P. and Coller, H. (2012) Functional Interactions Between microRNAs and RNA Binding Proteins. *MicroRNA e,* **1**(1), 70–79.

[129] Jacobsen, A., Wen, J., Marks, D. S., and Krogh, A. (2010) Signatures of RNA binding proteins globally coupled to effective microRNA target sites. *Genome Research,* **20**(8), 1010–1019.

[130] Incarnato, D., Neri, F., Diamanti, D., and Oliviero, S. (2013) MREdictor: a two-step dynamic interaction model that accounts for mRNA accessibility and Pumilio binding accurately predicts microRNA targets. *Nucleic Acids Research,* **41**(18), 8421–8433.

[131] Jiang, P., Singh, M., and Coller, H. A. (2013) Computational Assessment of the Cooperativity between RNA Binding Proteins and MicroRNAs in Transcript Decay. *PLoS Comput Biol,* **9**(5), e1003075.

[132] Nam, J.-W., Rissland, O. S., Koppstein, D., Abreu-Goodger, C., Jan, C. H., Agarwal, V., Yildirim, M. A., Rodriguez, A., and Bartel, D. P. (2014) Global analyses of the effect of different cellular contexts on microRNA targeting. *Molecular Cell,* **53**(6), 1031–1043.

[133] Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature,* **505**(7485), 701–705.

[134] Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., and Chang, H. Y. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature,* **519**(7544), 486–490.

[135] Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R. C., Snyder, M. P., Segal, E., and Chang, H. Y. (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature,* **505**(7485), 706–709.

[136] Gaidatzis, D., van Nimwegen, E., Hausser, J., and Zavolan, M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC bioinformatics,* **8**, 69.

[137] He, F. and Jacobson, A. (2015) Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annual Review of Genetics,* **49**(1), 339–366.

[138] Toma, K. G., Rebbapragada, I., Durand, S., and Lykke-Andersen, J. (2015) Identification of elements in human long 3 UTRs that inhibit nonsense-mediated decay. *RNA,* **21**(5), 887–897.

[139] Ge, Z., Quek, B. L., Beemon, K. L., and Hogg, J. R. (2016) Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *eLife,* **5**, e11155.

[140] Hoshino, S.-i. (2012) Mechanism of the initiation of mRNA decay: role of eRF3 family G proteins. *Wiley Interdisciplinary Reviews: RNA,* **3**(6), 743–757.

[141] Shi, Y. (2012) Alternative polyadenylation: new insights from global analyses. *RNA,* **18**(12), 2105–2117.

[142] Elkon, R., Ugalde, A. P., and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics,* **14**(7), 496–506.

[143] Spies, N., Burge, C. B., and Bartel, D. P. (2013) 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Research,* **23**(12), 2078–2090.

[144] Bhler, M., Steiner, S., Mohn, F., Paillusson, A., and Mhlemann, O. (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3′ UTR length. *Nature Structural & Molecular Biology,* **13**(5), 462–464.

[145] Eberle, A. B., Stalder, L., Mathys, H., Orozco, R. Z., and Mhlemann, O. (2008) Post-transcriptional Gene Regulation by Spatial Rearrangement of the 3 Untranslated Region. *PLoS Biol,* **6**(4), e92.

[146] Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N. H. (2014) 3 UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons. *Cell Reports,* **9**(2), 555–568.

[147] Hogg, J. R. and Goff, S. P. (2010) Upf1 Senses 3UTR Length to Potentiate mRNA Decay. *Cell,* **143**(3), 379–389.

[148] Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008) A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS biology,* **6**(4), e111.

[149] Yang, E., Nimwegen, E. v., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J. E. (2003) Decay Rates of Human mRNAs: Correlation With Functional Characteristics and Sequence Attributes. *Genome Research,* **13**(8), 1863–1872.

[150] Harrison, P. F., Powell, D. R., Clancy, J. L., Preiss, T., Boag, P. R., Traven, A., Seemann, T., and Beilharz, T. H. (2015) PAT-seq: a method to study the integration of 3′-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA,* **21**(8), 1502–1510.

[151] Preiss, T., Muckenthaler, M., and Hentze, M. W. (1998) Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA,* **4**(11), 1321–1331.

[152] Beilharz, T. H. and Preiss, T. (2007) Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA,* **13**(7), 982–997.

[153] Jalkanen, A. L., Coleman, S. J., and Wilusz, J. (2014) Determinants and implications of mRNA poly(A) tail size–does this protein make my tail look big?. *Seminars in Cell & Developmental Biology,* **34**, 24–32.

[154] Kojima, S., Sher-Chen, E. L., and Green, C. B. (2012) Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes & Development,* **26**(24), 2724–2736.

[155] Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., and Bartel, D. P. (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature,* **508**(7494), 66–71.

[156] Chang, H., Lim, J., Ha, M., and Kim, V. N. (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. *Molecular Cell,* **53**(6), 1044–1052.

[157] Lim, J., Ha, M., Chang, H., Kwon, S. C., Simanshu, D. K., Patel, D. J., and Kim, V. N. (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell,* **159**(6), 1365–1376.

[158] Barrett, L. W., Fletcher, S., and Wilton, S. D. (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences,* **69**(21), 3613–3634.

[159] Bartel, D. P. (2009) MicroRNA Target Recognition and Regulatory Functions. *Cell,* **136**(2), 215–233.

[160] Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. (2010) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research,* p. gkq1069.

[161] Fox, M., Urano, J., and Reijo Pera, R. A. (2005) Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. *Genomics,* **85**(1), 92–105.

[162] Didiano, D. and Hobert, O. (2008) Molecular architecture of a miRNA-regulated 3 UTR. *RNA,* **14**(7), 1297–1317.

[163] Kim, B. C., Lee, H. C., Lee, J.-J., Choi, C.-M., Kim, D.-K., Lee, J. C., Ko, Y.-G., and Lee, J.-S. (2012) Wig1 prevents cellular senescence by regulating p21 mRNA decay through control of RISC recruitment. *The EMBO Journal,* **31**(22), 4289–4303.

[164] Lveill, N., Elkon, R., Davalos, V., Manoharan, V., Hollingworth, D., Vrielink, J. O., le Sage, C., Melo, C. A., Horlings, H. M., Wesseling, J., Ule, J., Esteller, M., Ramos, A., and Agami, R. (2011) Selective inhibition of microRNA accessibility by RBM38 is required for p53 activity. *Nature Communications,* **2**, 513.

[165] Borrmann, L., Wilkening, S., and Bullerdiek, J. (2001) The expression of HMGA genes is regulated by their 3UTR. *Oncogene,* **20**(33), 4537–4541.

[166] Morishita, A., Zaidi, M. R., Mitoro, A., Sankarasharma, D., Szabolcs, M., Okada, Y., D'Armiento, J., and Chada, K. (2013) HMGA2 Is a Driver of Tumor Metastasis. *Cancer Research,* **73**(14), 4289–4299.

[167] Lee, Y. S. and Dutta, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes & Development,* **21**(9), 1025–1030.

[168] Mayr, C., Hemann, M. T., and Bartel, D. P. (2007) Disrupting the Pairing Between let-7 and Hmga2 Enhances Oncogenic Transformation. *Science,* **315**(5818), 1576–1579.

[169] Jnson, L., Christiansen, J., Hansen, T. V. O., Vikes, J., Yamamoto, Y., and Nielsen, F. C. (2014) IMP3 RNP Safe Houses Prevent miRNA-Directed HMGA2 mRNA Decay in Cancer and Development. *Cell Reports,* **7**(2), 539–551.

[170] Khaziapoul, S., Pearson, M. J., Pryme, I. F., Stern, B., and Hesketh, J. E. (2012) CUG binding protein 1 binds to a specific region within the human albumin 3 untranslated region. *Biochemical and Biophysical Research Communications,* **426**(4), 539–543.

[171] Diab, T., Hanoun, N., Bureau, C., Christol, C., Buscail, L., Cordelier, P., and Torrisani, J. (2013) The Role of the 3′ Untranslated Region in the Post-Transcriptional

Regulation of KLF6 Gene Expression in Hepatocellular Carcinoma. *Cancers,* **6**(1), 28–41.

[172] Melanson, B. D., Cabrita, M. A., Bose, R., Hamill, J. D., Pan, E., Brochu, C., Marcellus, K. A., Zhao, T. T., Holcik, M., and McKay, B. C. (2013) A novel cis-acting element from the 3UTR of DNA damage-binding protein 2 mRNA links transcriptional and post-transcriptional regulation of gene expression. *Nucleic Acids Research,* **41**(11), 5692–5703.

[173] Wirsing, A., Senkel, S., Klein-Hitpass, L., and Ryffel, G. U. (2011) A Systematic Analysis of the 3UTR of HNF4A mRNA Reveals an Interplay of Regulatory Elements Including miRNA Target Sites. *PLoS ONE,* **6**(11), e27438.

[174] Tanguay, R. L. and Gallie, D. R. (1996) Translational efficiency is regulated by the length of the 3' untranslated region.. *Molecular and Cellular Biology,* **16**(1), 146–156.

[175] Nicholson, P., Yepiskoposyan, H., Metze, S., Orozco, R. Z., Kleinschmidt, N., and Mhlemann, O. (2009) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cellular and Molecular Life Sciences,* **67**(5), 677–700.

[176] Zovoilis, A., Nolte, J., Drusenheimer, N., Zechner, U., Hada, H., Guan, K., Hasenfuss, G., Nayernia, K., and Engel, W. (2008) Multipotent adult germline stem cells and embryonic stem cells have similar microRNA profiles. *Molecular Human Reproduction,* **14**(9), 521–529.

[177] Wang, Z., Petersen, K., Weaver, M. S., and Magnuson, N. S. (2001) cDNA cloning, sequencing and characterization of bovine pim-1. *Veterinary Immunology and Immunopathology,* **78**(2), 177–195.

[178] Fan, X. C. and Steitz, J. A. (1998) Overexpression of HuR, a nuclearcytoplasmic shuttling protein, increases the in vivo stability of AREcontaining mRNAs. *The EMBO Journal,* **17**(12), 3448–3460.

[179] Peng, S. S.-Y., Chen, C.-Y. A., Xu, N., and Shyu, A.-B. (1998) RNA stabilization by the AUrich element binding protein, HuR, an ELAV protein. *The EMBO Journal,* **17**(12), 3461–3470.

[180] Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods,* **8**(7), 559–564.

[181] Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin,

S.-C., Gram, H., and Han, J. (2005) Involvement of MicroRNA in AU-Rich Element-Mediated mRNA Instability. *Cell*, **120**(5), 623–634.

[182] Pullmann, R., Kim, H. H., Abdelmohsen, K., Lal, A., Martindale, J. L., Yang, X., and Gorospe, M. (2007) Analysis of Turnover and Translation Regulatory RNA-Binding Protein Expression through Binding to Cognate mRNAs. *Molecular and Cellular Biology*, **27**(18), 6265–6278.

[183] Nei, M. and Li, W. H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases.. *Proceedings of the National Academy of Sciences of the United States of America*, **76**(10), 5269–5273.

[184] Root, D. E., Hacohen, N., Hahn, W. C., Lander, E. S., and Sabatini, D. M. (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nature Methods*, **3**(9), 715–719.

[185] Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., and Mayr, C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & Development*, **27**(21), 2380–2396.

[186] Chen, C.-Y., Chen, S.-T., Juan, H.-F., and Huang, H.-C. (2012) Lengthening of 3'UTR increases with morphological complexity in animal evolution. *Bioinformatics*, **28**(24), 3178–3181.

[187] Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. H. (2009) Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. *DNA Research*, **16**(1), 45–58.

[188] Popp, M. W.-L. and Maquat, L. E. (2013) Organizing Principles of Mammalian Nonsense-Mediated mRNA Decay. *Annual Review of Genetics*, **47**(1), 139–165.

[189] Schweingruber, C., Rufener, S. C., Znd, D., Yamashita, A., and Mhlemann, O. (2013) Nonsense-mediated mRNA decay Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1829**(67), 612–623.

[190] Peccarelli, M. and Kebaara, B. W. (2014) Regulation of Natural mRNAs by the Nonsense-Mediated mRNA Decay Pathway. *Eukaryotic Cell*, **13**(9), 1126–1135.

[191] Wittmann, J., Hol, E. M., and Jck, H.-M. (2006) hUPF2 Silencing Identifies Physiologic Substrates of Mammalian Nonsense-Mediated mRNA Decay. *Molecular and Cellular Biology*, **26**(4), 1272–1287.

[192] Kebaara, B. W. and Atkin, A. L. (2009) Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Research,* **37**(9), 2771–2778.

[193] Hurt, J. A., Robertson, A. D., and Burge, C. B. (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Research,* **23**(10), 1636–1650.

[194] Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., and Koller, D. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Molecular Systems Biology,* **10**, 770.

[195] Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature,* **446**(7138), 926–929.

[196] Pestka, S. (1971) Inhibitors of Ribosome Functions. *Annual Review of Microbiology,* **25**(1), 487–562.

[197] Ochs, M. J., Sorg, B. L., Pufahl, L., Grez, M., Suess, B., and Steinhilber, D. (2012) Post-Transcriptional Regulation of 5-Lipoxygenase mRNA Expression via Alternative Splicing and Nonsense-Mediated mRNA Decay. *PLOS ONE,* **7**(2), e31363.

[198] Yan, Y.-B. (2014) Deadenylation: enzymes, regulation, and functional implications. *Wiley Interdisciplinary Reviews: RNA,* **5**(3), 421–443.

[199] BehmAnsmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007) A conserved role for cytoplasmic poly(A)binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *The EMBO Journal,* **26**(6), 1591–1601.

[200] Choe, J., Cho, H., Chi, S.-G., and Kim, Y. K. (2011) Ago2/miRISC-mediated inhibition of CBP80/20-dependent translation and thereby abrogation of nonsense-mediated mRNA decay require the cap-associating activity of Ago2. *FEBS letters,* **585**(17), 2682–2687.

[201] Pesole, G., Liuni, S., Grillo, G., and Saccone, C. (1997) Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene,* **205**(12), 95–102.